

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-39363

(P2003-39363A)

(43) 公開日 平成15年2月13日 (2003.2.13)

(51) Int.Cl.<sup>7</sup>

識別記号

F I

テーマコード(参考)

B 2 5 J 13/00

B 2 5 J 13/00

Z 2 C 1 5 0

A 6 3 H 11/00

A 6 3 H 11/00

Z 3 C 0 0 7

B 2 5 J 5/00

B 2 5 J 5/00

C

審査請求 未請求 請求項の数32 O L (全 20 頁)

(21) 出願番号 特願2002-127374(P2002-127374)

(22) 出願日 平成14年4月26日 (2002.4.26)

(31) 優先権主張番号 0 1 4 0 1 1 2 7 . 4

(32) 優先日 平成13年4月30日 (2001.4.30)

(33) 優先権主張国 欧州特許庁 (E P)

(71) 出願人 599072932

ソニー フランス エスアー

フランス国 75831 パリ市 セデックス

17番フロレアル通り 15番地

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(74) 代理人 100067736

弁理士 小池 晃 (外2名)

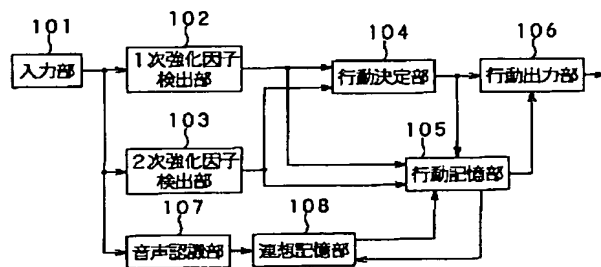
最終頁に続く

(54) 【発明の名称】 ロボット装置、ロボット装置の行動学習方法、ロボット装置の行動学習プログラム、及びプログラム記録媒体

(57) 【要約】

【課題】 複雑で連続的な行動等を効率的に学習する。

【解決手段】 ロボット装置では、行動決定部104で確率的に決定された行動に基づいて、行動出力部106がその行動を実行させるコマンドを出力する。所望の行動に近づくと、訓練者は、ロボット装置に2次強化因子を与え、所望の行動に近づくための標識とする。2次強化因子が検出されると、行動決定部104は、所望の行動への遷移確率を変化させる。最終的にロボット装置が所望の行動を行うと、訓練者は、ロボット装置に1次強化因子を与える。所望の行動に至る一連の行動とコマンドとを関連づける場合には、行動記憶部105から読み出された一連の行動と音声認識部107で認識されたコマンドとが、連想記憶部108にて連想記憶される。



## 【特許請求の範囲】

【請求項 1】 動作部を制御して行動を出現させるロボット装置であって、

行動の出現傾向に基づいて、一の行動を決定する行動決定手段と、

上記行動決定手段が決定した行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力手段と、

外部情報を入力する入力手段と、

上記入力手段が入力した上記外部情報から予め設定された 1 次強化因子を検出する 1 次強化因子検出手段と、

上記入力手段が入力した上記外部情報から 2 次強化因子を検出する 2 次強化因子検出手段と、

上記行動決定手段が決定した行動を記憶する行動記憶手段とを備え、

上記行動決定手段は、上記 2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向を高め、

上記行動記憶手段は、上記 1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動を記憶することを特徴とするロボット装置。

【請求項 2】 上記 2 次強化因子が検出された行動は、上記 1 次強化因子が検出された行動に至るまでの行動であることを特徴とする請求項 1 記載のロボット装置。

【請求項 3】 上記行動記憶手段は、上記 1 次強化因子が検出された行動を、当該行動前に上記 2 次強化因子が検出された行動との関連性のもとに記憶することを特徴とする請求項 1 記載のロボット装置。

【請求項 4】 上記行動記憶手段は、上記 1 次強化因子が検出された行動までに上記 2 次強化因子が検出された行動をシーケンス行動として記憶することを特徴とする請求項 3 記載のロボット装置。

【請求項 5】 上記行動決定手段は、上記シーケンス行動に含まれる各行動の出現傾向を高めることを特徴とする請求項 4 記載のロボット装置。

【請求項 6】 上記 1 次強化因子が検出されると、上記シーケンス行動を順に再現するようになされており、上記行動出力手段は、上記シーケンス行動に含まれる各行動を順に出現させ、その結果、上記 1 次強化因子が検出された場合には、上記行動記憶手段は、上記シーケンス行動を確定して記憶し、上記 1 次強化因子が検出されなかった場合には、上記行動出力手段は、上記 1 次強化因子が検出されるまで上記シーケンス行動に含まれる各行動の数を減らして順に出現させることを特徴とする請求項 4 記載のロボット装置。

【請求項 7】 上記入力手段が入力した所定の外部情報と上記行動記憶手段が記憶した上記 1 次強化因子が検出された行動とを関連づけて連想記憶する連想記憶手段を備えることを特徴とする請求項 1 記載のロボット装置。

【請求項 8】 上記入力手段が入力した音声情報を音声

処理して所定の言語として認識する音声認識手段を備え、

上記連想記憶手段は、上記音声処理により認識された所定のコマンドと上記行動記憶手段が記憶した上記 1 次強化因子が検出された行動とを関連づけて連想記憶することを特徴とする請求項 7 記載のロボット装置。

【請求項 9】 上記所定のコマンドは、上記 1 次強化因子が検出された後に入力されたものであることを特徴とする請求項 8 記載のロボット装置。

【請求項 10】 上記入力手段が入力した音声情報を音声処理して所定の言語として認識する音声認識手段と、上記音声処理により認識された上記シーケンス行動に関わる所定のコマンドと上記行動記憶手段が記憶した上記シーケンス行動とを関連づけて連想記憶する連想記憶手段とを備えることを特徴とする請求項 4 記載のロボット装置。

【請求項 11】 行動を学習するための行動学習モードを有し、

上記行動記憶手段は、上記行動学習モードにおいて上記 1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動を記憶することを特徴とする請求項 8 記載のロボット装置。

【請求項 12】 上記所定のコマンドは、上記行動学習モードとされた直後に入力されたものであることを特徴とする請求項 11 記載のロボット装置。

【請求項 13】 2 次強化因子を設定するための 2 次強化因子設定モードを有し、

上記 2 次強化因子設定モードにおいて、任意の外部情報が入力された後に上記 1 次強化因子が検出されると、当該外部情報が上記 2 次強化因子とされることを特徴とする請求項 1 記載のロボット装置。

【請求項 14】 任意の外部情報が入力された後に上記 1 次強化因子が検出されたことが所定回数以上連続すると、当該外部情報が上記 2 次強化因子とされることを特徴とする請求項 1 記載のロボット装置。

【請求項 15】 上記 1 次強化因子及び上記 2 次強化因子は、上記入力手段が入力した音声情報であることを特徴とする請求項 1 記載のロボット装置。

【請求項 16】 動作部を制御して行動を出現させるロボット装置の行動学習方法であって、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、

上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、

外部情報を入力する入力手段が入力した上記外部情報から予め設定された 1 次強化因子を検出する 1 次強化因子検出工程と、

上記入力手段が入力した上記外部情報から 2 次強化因子を検出する 2 次強化因子検出工程と、

上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、

上記行動決定工程では、上記 2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向が高められ、

上記行動記憶工程では、上記 1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶されることを特徴とするロボット装置の行動学習方法。

【請求項 17】 上記 2 次強化因子が検出された行動は、上記 1 次強化因子が検出された行動に至るまでの行動であることを特徴とする請求項 16 記載のロボット装置の行動学習方法。

【請求項 18】 上記行動記憶工程では、上記 1 次強化因子が検出された行動が、当該行動前に上記 2 次強化因子が検出された行動との関連性のもとに記憶されることを特徴とする請求項 17 記載のロボット装置の行動学習方法。

【請求項 19】 上記行動記憶工程では、上記 1 次強化因子が検出された行動までに上記 2 次強化因子が検出された行動がシーケンス行動として記憶されることを特徴とする請求項 18 記載のロボット装置の行動学習方法。

【請求項 20】 上記行動決定工程では、上記シーケンス行動に含まれる各行動の出現傾向が高められることを特徴とする請求項 19 記載のロボット装置の行動学習方法。

【請求項 21】 上記 1 次強化因子が検出されると、上記シーケンス行動を順に再現するようになされており、上記行動出力工程では、上記シーケンス行動に含まれる各行動が順に出現され、その結果、上記 1 次強化因子が検出された場合には、上記行動記憶工程では、上記シーケンス行動を確定して記憶し、上記 1 次強化因子が検出されなかった場合には、上記行動出力工程では、上記 1 次強化因子が検出されるまで上記シーケンス行動に含まれる各行動の数が減らされて順に出現されることを特徴とする請求項 19 記載のロボット装置の行動学習方法。

【請求項 22】 上記入力手段が入力した所定の外部情報と上記行動記憶工程で記憶された上記 1 次強化因子が検出された行動とを関連づけて連想記憶する連想記憶工程を有することを特徴とする請求項 16 記載のロボット装置の行動学習方法。

【請求項 23】 上記入力手段が入力した音声情報を音声処理して所定の言語として認識する音声認識工程を有し、  
上記連想記憶工程では、上記音声処理により認識された所定のコマンドと上記行動記憶工程で記憶された上記 1 次強化因子が検出された行動とを関連づけて連想記憶することを特徴とする請求項 22 記載のロボット装置の行動学習方法。

【請求項 24】 上記所定のコマンドは、上記 1 次強化

因子が検出された後に入力されたものであることを特徴とする請求項 23 記載のロボット装置の行動学習方法。

【請求項 25】 上記入力手段が入力した音声情報を音声処理して所定の言語として認識する音声認識工程と、上記音声処理により認識された上記シーケンス行動に関わる所定のコマンドと上記行動記憶工程で記憶された上記シーケンス行動とを関連づけて連想記憶する連想記憶工程とを有することを特徴とする請求項 19 記載のロボット装置の行動学習方法。

10 【請求項 26】 行動を学習するための行動学習モードを有し、

上記行動記憶工程では、上記行動学習モードにおいて上記 1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶されることを特徴とする請求項 23 記載のロボット装置の行動学習方法。

【請求項 27】 上記所定のコマンドは、上記行動学習モードとされた直後に入力されたものであることを特徴とする請求項 26 記載のロボット装置の行動学習方法。

20 【請求項 28】 2 次強化因子を設定するための 2 次強化因子設定モードを有し、

上記 2 次強化因子設定モードにおいて、任意の外部情報が入力された後に上記 1 次強化因子が検出されると、当該外部情報が上記 2 次強化因子とされることを特徴とする請求項 16 記載のロボット装置の行動学習方法。

【請求項 29】 任意の外部情報が入力された後に上記 1 次強化因子が検出されたことが所定回数以上連続すると、当該外部情報が上記 2 次強化因子とされることを特徴とする請求項 16 記載のロボット装置の行動学習方法。

30 【請求項 30】 上記 1 次強化因子及び上記 2 次強化因子は、上記入力手段が入力した音声情報であることを特徴とする請求項 16 記載のロボット装置の行動学習方法。

【請求項 31】 動作部を制御して行動を出現させるロボット装置の行動学習プログラムであって、  
行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、

上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、

外部情報を入力する入力手段が入力した上記外部情報から予め設定された 1 次強化因子を検出する 1 次強化因子検出工程と、

上記入力手段が入力した上記外部情報から 2 次強化因子を検出する 2 次強化因子検出工程と、

上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、

上記行動決定工程では、上記 2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向が高められ、

50

上記行動記憶工程では、上記1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記憶されることを特徴とするロボット装置の行動学習プログラム。

【請求項32】 動作部を制御して行動を出現させるロボット装置の行動学習プログラムが記録されたコンピュータ制御可能なプログラム記録媒体であって、上記ロボット装置の行動学習プログラムは、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、

上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、

外部情報を入力する入力手段が入力した上記外部情報から予め設定された1次強化因子を検出する1次強化因子検出工程と、

上記入力手段が入力した上記外部情報から2次強化因子を検出する2次強化因子検出工程と、

上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、

上記行動決定工程では、上記2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、

上記行動記憶工程では、上記1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記憶されることを特徴とするロボット装置の行動学習プログラムが記録されたプログラム記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、自律的に行動するロボット装置、そのようなロボット装置が所望の行動を学習するロボット装置の行動学習方法、ロボット装置の行動学習プログラム、及び当該ロボット装置の行動学習プログラムの記録されたプログラム記録媒体に関する。

【0002】

【従来の技術】従来より、犬等の動物を飼い慣らすために、調教が行われている。例えば、犬の飼い主は、「お手」や「お座り」等の行動を犬に対して教示する。しかし、多くの犬の飼い主は、犬を所望の位置につかせようとしながら、同時に命令するという間違いを犯している。例えば、訓練者は、「お座り」という言葉を繰り返しながら、犬のお尻を地面に向かって押下げる。この方法が成功しないのは、以下のような理由からである。

【0003】(1) 動物は、訓練者の言葉が学習行動のいずれかに注意を払う選択を強いられる。

【0004】(2) 命令を複数回繰り返すため、動物はそのうちのどの部分が行動と関連しているのかわからない。

【0005】(3) 行動する前に命令される場合が非常に多い。例えば、動物がまだ立っているうちに「お座

り」と命令してしまうため、「お座り」を座った状態と関連させることができない。

【0006】このような理由により、多くの訓練者は命令と行動とを別々に教える。実際には最初に行動を教え、次に命令を教える。

【0007】動物に行動を教えるための方法には、代表的なものとして、モデリング (Modelling)、ルアーリング (Ruring)、キャプチャリング (Capturing)、イミテーション (Imitating) 等があり、それぞれ図14に示すような特徴を有する。以下、簡単に説明する。

【0008】モデリングは、多くの犬の飼い主が試行する方法であるが、専門家である調教師が行うことは決していない。この方法は、動物を所望のポジションへ物理的に操り、そのポジションにつくことができた褒めてあげるというものである。この方法では、動物は受身のままである。そのため、多くの場合において行動学習の成果が上がらない。また、静止ポジションへ達する以上の複雑な動きをこの方法により実現するのは容易ではない。

【0009】ルアーリングは、動物との物理的な接触がない点を除いてモデリングと類似している。訓練者は、玩具や食べ物を動物の鼻の前に置き、これを使って動物を所望のポジションに誘導することができる。この方法は、実際の動物の場合には良い成果が得られるが、その使用は、あるポジションや非常に単純な動作を教えるときに限定される。

【0010】キャプチャリングは、上述したモデリングやルアーリングとは異なり、動物の自発的な行動を利用したものである。例えば、訓練者は、動物が所望のポジションについている、又は正確な行動をとっていると認識する毎に、ポジティブな強化因子を動物に与える。この方法についても、実際の動物の場合には良い成果が得られるが、訓練者は、動物が所望のポジションにつくまで待つ必要がある。例えば、「お座り」を教えるときには、訓練者は、動物が自発的に座るのを待つ必要がある。また、この方法は、動物の自発的な行動を利用したものであるため、連続的な動作や珍しい動作を教えることが困難である。

【0011】イミテーションは、動物に訓練者と同じ動作を模倣させるものであるが、訓練者がイミテーションに基づいた方法を用いることは滅多にない。この方法では、連続的な動作や珍しい動作についても教えることができるが、この方法は、霊長類や海洋哺乳動物等の高等動物についてしか確認されていない。

【0012】

【発明が解決しようとする課題】ところで、近年、動物を模した形状とされたロボット装置、いわゆるペットロボットが提案されている。このようなロボット装置は、一般家庭において飼育される犬や猫に似た形状を有し、ユーザ (飼い主) からの「叩く」や「撫でる」といった

働きかけや、周囲の環境等に応じて自律的に行動する。  
例えば、自律的な行動として、実際の動物と同様に、  
「吠える」、「寝る」等といった行動をする。

【0013】このようなロボット装置が実際の動物のような行動に、より近づくことができれば、ロボット装置の生物感が一層増し、ユーザ（飼い主）は、ロボット装置に一層の親近感や満足感を感じる。これにより、ロボット装置のアミューズメント性が向上する。

【0014】例えば、実際の動物に対するのと同様に、ユーザ（飼い主）が自らの所有するロボット装置に対し  
て所望の動作を教示し、ロボット装置を“飼い慣らす”  
ことができれば、ユーザ（飼い主）は、ロボット装置に  
一層の親近感や満足感を感じるようになると考えられ  
る。

【0015】しかし、上述したような実際の動物に行動を教えるための方法を、自律的に行動するロボット装置に適用するには、種々の問題が存在する。

【0016】例えば、上述したモデリングは、産業界においてロボット装置にポジションを教えるために多く用いられてきたが、ロボット装置が自律的に行動し、常時  
活発になると問題が生じてくる。すなわち、ロボット装置は、例えば、訓練者が背中を押していることを感知することができ、また、プログラムされていれば座る判断を下すことができる。しかし、静止ポジションへ達する以上の複雑な動きをこの方法により実現するのは容易ではないという問題がある。

【0017】また、上述したルアリングは、ロボット工学ではあまり使用されていない。ロボット装置が自動的に所定のもの、例えば赤いものに興味を示すようにプログラムされていれば、訓練者は、この特性を利用してロボット装置を所望のポジションにつかせることができる。しかし、これも使用範囲が限られるという問題がある。

【0018】上述したキャプチャリングは、この方法を用いて、ロボット装置が信号（例えば言葉）と関連した所望の行動をとる毎に、その信号を後から発信するというモデルを作成することができる。しかし、ロボット装置に「お座り」のような言葉を教えるには、訓練者は、ロボット装置が自発的に座るのを待つ必要があるという問題がある。また、名前をつけることができる動作の数が多過ぎるという問題や、動物の自発的な行動を利用したものであるため、連続的な動作や珍しい動作を教えることが困難であるといった問題がある。

【0019】上述したイミテーションは、例えば、「Y. Kuniyoshi et al., "Learning by watching: Extracting reusable task knowledge from visual observation of human performance.", IEEE Transactions on Robotics and Automation, 10(6):799-822, 1994」等に記載されているように、ロボット工学において、いくつかの研究グループが取り組んできた。この方法では、原理的に、連

続した動作や非常に珍しい動作を教えることができるが、多大な量の計算能力を必要とするため、この方法を既存の自律的に行動するロボット装置に適用することは困難である。

【0020】本発明は、このような従来の実情に鑑みて提案されたものであり、複雑で連続的な行動等を修得することのできるロボット装置、このようなロボット装置の行動学習方法、ロボット装置の行動学習プログラム、及び当該ロボット装置の行動学習プログラムの記録されたプログラム記録媒体を提供することを目的とする。

【0021】

【課題を解決するための手段】上述した目的を達成するために、本発明に係るロボット装置は、動作部を制御して行動を出現させるロボット装置であって、行動の出現傾向に基づいて、一の行動を決定する行動決定手段と、上記行動決定手段が決定した行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力手段と、外部情報を入力する入力手段と、上記入力手段が入力した上記外部情報から予め設定された1次強化因子を検出する1次強化因子検出手段と、上記入力手段が入力した上記外部情報から2次強化因子を検出する2次強化因子検出手段と、上記行動決定手段が決定した行動を記憶する行動記憶手段とを備え、上記行動決定手段は、上記2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向を高め、上記行動記憶手段は、上記1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動を記憶することを特徴としている。

【0022】ここで、2次強化因子が検出された行動は、1次強化因子が検出された行動に至るまでの行動である。

【0023】このようなロボット装置は、行動の学習時に、2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向を高め、1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動を記憶する。

【0024】また、上述した目的を達成するために、本発明に係るロボット装置の行動学習方法は、動作部を制御して行動を出現させるロボット装置の行動学習方法であって、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、外部情報を入力する入力手段が入力した上記外部情報から予め設定された1次強化因子を検出する1次強化因子検出工程と、上記入力手段が入力した上記外部情報から2次強化因子を検出する2次強化因子検出工程と、上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、上記行動決定工程では、上記2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、

上記行動記憶工程では、上記 1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶されることを特徴としている。

【0025】ここで、2 次強化因子が検出された行動は、1 次強化因子が検出された行動に至るまでの行動である。

【0026】このようなロボット装置の行動学習方法では、行動の学習時に、2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向が高められ、1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶される。

【0027】また、上述した目的を達成するために、本発明に係るロボット装置の行動学習プログラムは、動作部を制御して行動を出現させるロボット装置の行動学習プログラムであって、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、外部情報を入力する入力手段が入力した上記外部情報から予め設定された 1 次強化因子を検出する 1 次強化因子検出工程と、上記入力手段が入力した上記外部情報から 2 次強化因子を検出する 2 次強化因子検出工程と、上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、上記行動決定工程では、上記 2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向が高められ、上記行動記憶工程では、上記 1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶されることを特徴としている。

【0028】ここで、2 次強化因子が検出された行動は、1 次強化因子が検出された行動に至るまでの行動である。

【0029】このようなロボット装置の行動学習プログラムでは、行動の学習時に、2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向が高められ、1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶される。

【0030】また、上述した目的を達成するために、本発明に係るプログラム記録媒体は、動作部を制御して行動を出現させるロボット装置の行動学習プログラムが記録されたコンピュータ制御可能なプログラム記録媒体であって、上記ロボット装置の行動学習プログラムは、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、外部情報を入力する入力手段が入力した上記外部情報から予め設定された 1 次強化因子を検出する 1 次強化因子検出工程と、上記入力手段が入力した上記外部情報から 2 次強化因子を検出する 2 次強化因子検出工程と、上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、上記行動決定工程で

は、上記 2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向が高められ、上記行動記憶工程では、上記 1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶されることを特徴とするロボット装置の行動学習プログラムが記録されたものである。

【0031】ここで、2 次強化因子が検出された行動は、1 次強化因子が検出された行動に至るまでの行動である。

【0032】このようなプログラム記録媒体に記録されているロボット装置の行動学習プログラムでは、行動の学習時に、2 次強化因子が検出される毎に、当該 2 次強化因子が検出された行動の出現傾向が高められ、1 次強化因子が検出されると、少なくとも当該 1 次強化因子が検出された行動が記憶される。

【0033】

【発明の実施の形態】以下、本発明を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。この実施の形態は、本発明を、周囲の環境（外的要因）や内部の状態（内的要因）に応じて自律的に行動をするロボット装置に適用したものである。

【0034】実施の形態では、先ず、ロボット装置の構成について説明して、その後、ロボット装置における本発明の適用部分について詳細に説明する。

【0035】（1）本実施の形態におけるロボット装置の構成

図 1 に示すように、「犬」を模した形状のいわゆるペットロボットとされ、胴体部ユニット 2 の前後左右にそれぞれ脚部ユニット 3 A、3 B、3 C、3 D が連結されると共に、胴体部ユニット 2 の前端部及び後端部にそれぞれ頭部ユニット 4 及び尻尾部ユニット 5 が連結されて構成されている。

【0036】胴体部ユニット 2 には、図 2 に示すように、CPU（Central Processing Unit）10、DRAM（Dynamic Random Access Memory）11、フラッシュ ROM（Read Only Memory）12、PC（Personal Computer）カードインターフェース回路 13 及び信号処理回路 14 が内部バス 15 を介して相互に接続されることにより形成されたコントロール部 16 と、このロボット装置の動力源としてのバッテリー 17 とが収納されている。また、胴体部ユニット 2 には、ロボット装置 1 の向きや動きの加速度を検出するための角速度センサ 18 及び加速度センサ 19 なども収納されている。

【0037】また、頭部ユニット 4 には、外部の状況を撮像するための CCD（Charge Coupled Device）カメラ 20 と、ユーザからの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ 21 と、前方に位置する物体までの距離を測定するための距離センサ 22 と、外部音を集音するためのマイクロホン 23 と、鳴き声等の音声を出力するため

のスピーカ24と、ロボット装置1の「目」に相当するLED (Light Emitting Diode) (図示せず) などがそれぞれ所定位置に配置されている。

【0038】さらに、各脚部ユニット3A~3Dの関節部分や各脚部ユニット及び胴体部ユニット2の各連結部分、頭部ユニット4及び胴体部ユニット2の連結部分、並びに尻尾部ユニット5の尻尾5Aの連結部分などにはそれぞれ自由度数分のアクチュエータ251, 252, ...及びポテンシオメータ261, 262, ...が配設されている。例えば、アクチュエータ251, 252, ...は、サーボモータを構成として有している。サーボモータの駆動により、脚部ユニット3A~3Dが制御されて、目的の姿勢或いは動作に遷移する。

【0039】そして、これら角速度センサ18、加速度センサ19、タッチセンサ21、距離センサ22、マイクロホン23、スピーカ24及び各ポテンシオメータ261, 262, ...などの各種センサ並びにLED及び各アクチュエータ251, 252, ...は、それぞれ対応するハブ271~27nを介してコントロール部16の信号処理回路14と接続され、CCDカメラ20及びバッテリー17は、それぞれ信号処理回路14と直接接続されている。

【0040】信号処理回路14は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部バス15を介してDRAM11内の所定位置に順次格納する。また、信号処理回路14は、これと共にバッテリー17から供給されるバッテリー残量を表すバッテリー残量データを順次取り込み、これをDRAM11内の所定位置に格納する。

【0041】このようにしてDRAM11内に格納された各センサデータ、画像データ、音声データ及びバッテリー残量データは、この後CPU10がこのロボット装置1の動作制御を行う際に利用される。

【0042】實際上CPU10は、ロボット装置1の電源が投入された初期時、胴体部ユニット2の図示しないPCカードスロットに装填されたメモリカード28又はフラッシュROM12に格納された制御プログラムをPCカードインターフェース回路13を介して又は直接読み出し、これをDRAM11に格納する。

【0043】また、CPU10は、この後上述のように信号処理回路14よりDRAM11に順次格納される各センサデータ、画像データ、音声データ及びバッテリー残量データの基づいて自己及び周囲の状況や、ユーザからの指令及び働きかけの有無などを判断する。

【0044】さらに、CPU10は、この判断結果及びDRAM11に格納した制御プログラムに基づいて続く行動を決定すると共に、当該決定結果に基づいて必要なアクチュエータ251, 252, ...を駆動させることにより、頭部ユニット4を上下左右に振らせたり、尻尾部ユニット5の尻尾5Aを動かしたり、各脚部ユニッ

ト3A~3Dを駆動させて歩行させるなどの行動を行わせる。

【0045】また、この際CPU10は、必要に応じて音声データを生成し、これを信号処理回路14を介して音声信号としてスピーカ24に与えることにより、当該音声信号に基づく音声を外部に出力させたり、上述のLEDを点灯又は点滅させる。

【0046】このようにして、このロボット装置1においては、自己及び周囲の状況や、ユーザからの指令及び働きかけに応じて自律的に行動し得るようになされている。

【0047】(2) 制御プログラムのソフトウェア構成  
ここで、ロボット装置1における上述の制御プログラムのソフトウェア構成は、図3に示すようになる。この図3において、デバイス・ドライバ・レイヤ30は、この制御プログラムの最下位層に位置し、複数のデバイス・ドライバからなるデバイス・ドライバ・セット31から構成されている。この場合、各デバイス・ドライバは、CCDカメラ20(図2)や、タイマ等の通常のコンピュータで用いられるハードウェアに直接アクセスすることを許されたオブジェクトであり、対応するハードウェアからの割り込みを受けて処理を行う。

【0048】また、ロボティック・サーバ・オブジェクト32は、デバイス・ドライバ・レイヤ30の最下位層に位置し、例えば上述の各種センサやアクチュエータ251, 252, ...等のハードウェアにアクセスするためのインターフェースを提供するソフトウェア群でなるバーチャル・ロボット33と、電源の切り換えなどを管理するソフトウェア群でなるパワーマネージャ34と、他の種々のデバイス・ドライバを管理するソフトウェア群でなるデバイス・ドライバ・マネージャ35と、ロボット装置1の機構を管理するソフトウェア群でなるデザインド・ロボット36とから構成されている。

【0049】マネージャ・オブジェクト37は、オブジェクト・マネージャ38及びサービス・マネージャ39から構成されている。この場合、オブジェクト・マネージャ38は、ロボティック・サーバ・オブジェクト32、ミドル・ウェア・レイヤ40、及びアプリケーション・レイヤ41に含まれる各ソフトウェア群の起動や終了を管理するソフトウェア群であり、サービス・マネージャ39は、メモリカード28(図2)に格納されたコネクションファイルに記述されている各オブジェクト間の接続情報に基づいて各オブジェクトの接続を管理するソフトウェア群である。

【0050】ミドル・ウェア・レイヤ40は、ロボティック・サーバ・オブジェクト32の上位層に位置し、画像処理や音声処理などのこのロボット装置1の基本的な機能を提供するソフトウェア群から構成されている。また、アプリケーション・レイヤ41は、ミドル・ウェア・レイヤ40の上位層に位置し、当該ミドル・ウェア・

レイヤ40を構成する各ソフトウェア群によって処理された処理結果に基づいてロボット装置1の行動を決定するためのソフトウェア群から構成されている。

【0051】なお、ミドル・ウェア・レイヤ40及びアプリケーション・レイヤ41の具体的なソフトウェア構成を図4に示す。

【0052】ミドル・ウェア・レイヤ40においては、図4に示すように、騒音検出用、温度検出用、明るさ検出用、音階認識用、距離検出用、姿勢検出用、タッチセンサ用、動き検出用及び色認識用の各信号処理モジュール50～58並びに入力セマンティクスコンバータモジュール59などを有する認識系60と、出力セマンティクスコンバータモジュール68並びに姿勢管理用、トラッキング用、モーション再生用、歩行用、転倒復帰用、LED点灯用及び音再生用の各信号処理モジュール61～67などを有する出力系69とから構成されている。

【0053】認識系60の各信号処理モジュール50～58は、ロボティック・サーバ・オブジェクト32のバーチャル・ロボット33によりDRAM11（図2）から読み出される各センサデータや画像データ及び音声データのうちの対応するデータを取り込み、当該データに基づいて所定の処理を施して、処理結果を入力セマンティクスコンバータモジュール59に与える。ここで、例えば、バーチャル・ロボット33は、所定の通信規約によって、信号の授受或いは変換をする部分として構成されている。

【0054】入力セマンティクスコンバータモジュール59は、これら信号処理モジュール50～58から与えられる処理結果に基づいて、「うるさい」、「暑い」、「明るい」、「ボールを検出した」、「転倒を検出した」、「撫でられた」、「叩かれた」、「ドミソの音階が聞こえた」、「動く物体を検出した」又は「障害物を検出した」などの自己及び周囲の状況や、ユーザからの指令及び働きかけを認識し、認識結果をアプリケーション・レイヤ41（図3）に出力する。

【0055】アプリケーション・レイヤ41は、図5に示すように、行動モデルライブラリ70、行動切替モジュール71、学習モジュール72、感情モデル73及び本能モデル74の5つのモジュールから構成されている。

【0056】行動モデルライブラリ70には、図6に示すように、「バッテリー残量が少なくなった場合」、「転倒復帰する場合」、「障害物を回避する場合」、「感情を表現する場合」、「ボールを検出した場合」などの予め選択されたいくつかの条件項目にそれぞれ対応させて、それぞれ独立した行動モデル70<sub>1</sub>～70<sub>n</sub>が設けられている。

【0057】そして、これら行動モデル70<sub>1</sub>～70<sub>n</sub>は、それぞれ入力セマンティクスコンバータモジュール59から認識結果が与えられたときや、最後の認識結果

が与えられてから一定時間が経過したときなどに、必要に応じて後述のように感情モデル73に保持されている対応する情動のパラメータ値や、本能モデル74に保持されている対応する欲求のパラメータ値を参照しながら続く行動をそれぞれ決定し、決定結果を行動切替モジュール71に出力する。

【0058】なお、この実施の形態の場合、各行動モデル70<sub>1</sub>～70<sub>n</sub>は、次の行動を決定する手法として、図7に示すような1つのノード（状態）NODE<sub>o</sub>～NODE<sub>n</sub>から他のどのノードNODE<sub>o</sub>～NODE<sub>n</sub>に遷移するかを各ノードNODE<sub>o</sub>～NODE<sub>n</sub>の間を接続するアークARC<sub>1</sub>～ARC<sub>n</sub>に対してそれぞれ設定された遷移確率P<sub>1</sub>～P<sub>n</sub>に基づいて確率的に決定する有限確率オートマトンと呼ばれるアルゴリズムを用いる。

【0059】具体的に、各行動モデル70<sub>1</sub>～70<sub>n</sub>は、それぞれ自己の行動モデル70<sub>1</sub>～70<sub>n</sub>を形成するノードNODE<sub>o</sub>～NODE<sub>n</sub>にそれぞれ対応させて、これらノードNODE<sub>o</sub>～NODE<sub>n</sub>毎に図8に示すような状態遷移表80を有している。

【0060】この状態遷移表80では、そのノードにおいて遷移条件とする入力イベント（認識結果）が「入力イベント名」の列に優先順に列記され、その遷移条件についてのさらなる条件が「データ名」及び「データ範囲」の列における対応する行に記述されている。

【0061】したがって、図8の状態遷移表80で表されるノードNODE<sub>100</sub>では、「ボールを検出（BALL）」という認識結果が与えられた場合に、当該認識結果と共に与えられるそのボールの「大きさ（SIZE）」が「0から1000」の範囲であることや、「障害物を検出（OBSTACLE）」という認識結果が与えられた場合に、当該認識結果と共に与えられるその障害物までの「距離（DISTANCE）」が「0から100」の範囲であることが他のノードに遷移するための条件となっている。

【0062】また、このノードNODE<sub>100</sub>では、認識結果の入力がない場合においても、行動モデル70<sub>1</sub>～70<sub>n</sub>が周期的に参照する感情モデル73及び本能モデル74にそれぞれ保持された各情動及び各欲求のパラメータ値のうち、感情モデル73に保持された「喜び（JOY）」、「驚き（SURPRISE）」若しくは「悲しみ（SADNESS）」のいずれかのパラメータ値が「50から100」の範囲であるときは他のノードに遷移することができるようになっている。

【0063】また、状態遷移表80では、「他のノードへの遷移確率」の欄における「遷移先ノード」の行にそのノードNODE<sub>o</sub>～NODE<sub>n</sub>から遷移できるノード名が列記されていると共に、「入力イベント名」、「データ値」及び「データの範囲」の列に記述された全ての条件が揃ったときに遷移できる他の各ノードNODE<sub>o</sub>



～NODE<sub>n</sub>への遷移確率が「他のノードへの遷移確率」の欄内の対応する箇所にそれぞれ記述され、そのノードNODE<sub>o</sub>～NODE<sub>n</sub>に遷移する際に出力すべき行動が「他のノードへの遷移確率」の欄における「出力行動」の行に記述されている。なお、「他のノードへの遷移確率」の欄における各行の確率の和は100 [%]となっている。

【0064】したがって、図8の状態遷移表80で表されるノードNODE<sub>100</sub>では、例えば「ボールを検出(BALL)」し、そのボールの「大きさ(SIZE)」が「0から1000」の範囲であるという認識結果が与えられた場合には、「30 [%]」の確率で「ノードNODE<sub>120</sub> (node 120)」に遷移でき、そのとき「ACTION 1」の行動が出力されることとなる。

【0065】各行動モデル70<sub>1</sub>～70<sub>n</sub>は、それぞれこのような状態遷移表80として記述されたノードNODE<sub>o</sub>～NODE<sub>n</sub>がいくつも繋がるようにして構成されており、入力セマンティクスコンバータモジュール59から認識結果が与えられたときなどに、対応するノードNODE<sub>o</sub>～NODE<sub>n</sub>の状態遷移表80を利用して確率的に次の行動を決定し、決定結果を行動切替モジュール71に出力するようになされている。

【0066】図5に示す行動切替モジュール71は、行動モデルライブラリ70の各行動モデル70<sub>1</sub>～70<sub>n</sub>からそれぞれ出力される行動のうち、予め定められた優先順位の高い行動モデル70<sub>1</sub>～70<sub>n</sub>から出力された行動を選択し、当該行動を実行すべき旨のコマンド(以下、これを行動コマンドという。)をミドル・ウェア・レイヤ40の出力セマンティクスコンバータモジュール68に送出する。なお、この実施の形態においては、図6において下側に表記された行動モデル70<sub>1</sub>～70<sub>n</sub>ほど優先順位が高く設定されている。

【0067】また、行動切替モジュール71は、行動完了後に出力セマンティクスコンバータモジュール68から与えられる行動完了情報に基づいて、その行動が完了したことを学習モジュール72、感情モデル73及び本\*

$$E[t+1]=E[t]+k_e \times \Delta E[t]$$

【0073】なお、各認識結果や出力セマンティクスコンバータモジュール68からの通知が各情動のパラメータ値の変動量 $\Delta E[t]$ にどの程度影響を与えるかは予め決められており、例えば「叩かれた」といった認識結果は「怒り」の情動のパラメータ値の変動量 $\Delta E[t]$ に大きな影響を与え、「撫でられた」といった認識結果は「喜び」の情動のパラメータ値の変動量 $\Delta E[t]$ に大きな影響を与えるようになっている。

【0074】ここで、出力セマンティクスコンバータモジュール68からの通知とは、いわゆる行動のフィードバック情報(行動完了情報)であり、行動の出現結果の

\*能モデル74に通知する。

【0068】一方、学習モジュール72は、入力セマンティクスコンバータモジュール59から与えられる認識結果のうち、「叩かれた」や「撫でられた」など、使用者からの働きかけとして受けた教示の認識結果を入力する。

【0069】そして、学習モジュール72は、この認識結果及び行動切替モジュール71からの通知に基づいて、「叩かれた(叱られた)」ときにはその行動の発現確率を低下させ、「撫でられた(誉められた)」ときにはその行動の発現確率を上昇させるように、行動モデルライブラリ70における行動モデル70<sub>1</sub>～70<sub>n</sub>の対応する遷移確率を変更する。

【0070】他方、感情モデル73は、「喜び(joy)」、「悲しみ(sadness)」、「怒り(anger)」、「驚き(surprise)」、「嫌悪(disgust)」及び「恐れ(fear)」の合計6つの情動について、各情動ごとにその情動の強さを表すパラメータを保持している。そして、感情モデル73は、これら各情動のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール59から与えられる「叩かれた」及び「撫でられた」などの特定の認識結果と、経過時間及び行動切替モジュール71からの通知などに基づいて周期的に更新する。

【0071】具体的には、感情モデル73は、入力セマンティクスコンバータモジュール59から与えられる認識結果と、そのときのロボット装置1の行動と、前回更新してからの経過時間などに基づいて所定の演算式により算出されるそのときのその情動の変化量を $\Delta E$

[t]、現在のその情動のパラメータ値を $E[t]$ 、その情動の感度を表す係数を $k_e$ として、(1)式によって次の周期におけるその情動のパラメータ値 $E[t+1]$ を算出し、これを現在のその情動のパラメータ値 $E[t]$ と置き換えるようにしてその情動のパラメータ値を更新する。また、感情モデル73は、これと同様にして全ての情動のパラメータ値を更新する。

【0072】

$$\begin{aligned} & \text{【数1】} \\ & \dots (1) \end{aligned}$$

情報であり、感情モデル73は、このような情報によっても感情を変化させる。これは、例えば、「吠える」といった行動により怒りの感情レベルが下がるといったようなことである。なお、出力セマンティクスコンバータモジュール68からの通知は、上述した学習モジュール72にも入力されており、学習モジュール72は、その通知に基づいて行動モデル70<sub>1</sub>～70<sub>n</sub>の対応する遷移確率を変更する。

【0075】なお、行動結果のフィードバックは、行動切替モジュール71の出力(感情が付加された行動)によりなされるものであってもよい。

17

【0076】一方、本能モデル74は、「運動欲(exercise)」、「愛情欲(affection)」、「食欲(appetite)」及び「好奇心(curiosity)」の互いに独立した4つの欲求について、これら欲求ごとにその欲求の強さを表すパラメータを保持している。そして、本能モデル74は、これらの欲求のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール59から与えられる認識結果や、経過時間及び行動切替モジュール71からの通知などに基づいて周期的に更新する。

【0077】具体的には、本能モデル74は、「運動欲」、「愛情欲」及び「好奇心」については、認識結果、経過時間及び出力セマンティクスコンバータモジュール59から与えられる

$$I[k+1]=I[k]+k_i \times \Delta I[k]$$

【0079】なお、認識結果及び出力セマンティクスコンバータモジュール68からの通知などが各欲求のパラメータ値の変動量 $\Delta I[k]$ にどの程度影響を与えるかは予め決められており、例えば出力セマンティクスコンバータモジュール68からの通知は、「疲れ」のパラメータ値の変動量 $\Delta I[k]$ に大きな影響を与えるようになっている。

【0080】なお、本実施の形態においては、各情動及び各欲求(本能)のパラメータ値がそれぞれ0から100までの範囲で変動するように規制されており、また、係数 $k_e$ 、 $k_i$ の値も各情動及び各欲求ごとに個別に設定されている。

【0081】一方、ミドル・ウェア・レイヤ40の出力セマンティクスコンバータモジュール68は、図4に示すように、上述のようにして、アプリケーション・レイヤ41の行動切替モジュール71から与えられる「前進」、「喜ぶ」、「鳴く」又は「トラッキング(ボールを追いかける)」といった抽象的な行動コマンドを出力系69の対応する信号処理モジュール61~67に与える。

【0082】そしてこれら信号処理モジュール61~67は、行動コマンドが与えられると当該行動コマンドに基づいて、この行動を行うために対応するアクチュエータ251~25n(図2)に与えるべきサーボ指令値や、スピーカ24(図2)から出力する音の音声データ及び/又は「目」のLEDに与える駆動データを生成し、これらのデータをロボティック・サーバ・オブジェクト32のバーチャル・ロボット33及び信号処理回路14(図2)を順次介して対応するアクチュエータ251~25n又はスピーカ24又はLEDに順次送出する。

【0083】このようにしてロボット装置1においては、制御プログラムに基づいて、自己(内部)及び周囲(外部)の状況や、使用者からの指示及び働きかけに応じた自律的な行動を行うことができるようになされてい

18

\*ルール68からの通知などに基づいて所定の演算式により算出されるそのときのその欲求の変動量を $\Delta I[k]$ 、現在のその欲求のパラメータ値を $I[k]$ 、その欲求の感度を表す係数を $k_i$ として、所定期間で(2)式を用いて次の周期におけるその欲求のパラメータ値 $I[k+1]$ を算出し、この演算結果を現在のその欲求のパラメータ値 $I[k]$ と置き換えるようにしてその欲求のパラメータ値を更新する。また、本能モデル74は、これと同様にして「食欲」を除く各欲求のパラメータ値を更新する。

【0078】

【数2】

... (2)

る。

【0084】(3)本実施の形態で用いる行動教示方法本実施の形態では、ロボット装置1に行動を教示する方法として、シェーピング(Shaping)と呼ばれる方法を用いる。この方法では、行動の形成を行うために、行動を連続した達成可能な小さなレスポンスに分割し、最終的に所望の行動へ導く。すなわち、この方法は、適切な行動に向けて段階的に誘導するものである。各段階を実行するには種々の技術を用いることができるが、本実施の形態では、動物の訓練方法として普及しているクリッカートレーニングと呼ばれる方法を用いる。以下、このクリッカートレーニングについて説明する。

【0085】クリッカートレーニングは、「B.F. Skinner, "The Behavior of Organisms", Appleton Century Crofts, New York, NY., 1938」に記載されているB.F. Skinnerのオペラント条件付けの理論に基づいている。このクリッカートレーニングに基づいた犬の訓練方法は、1980年代に、行動主義心理学者のGary Wilkesと、イルカの訓練者であるKaren Pryorが共同で普及させた。クリッカートレーニングでは、従来よりイルカの訓練に使用されている笛の代わりに、金属製の小型玩具であるクリッカーを使用している。

【0086】クリッカーを押すと短く鋭い音を発する。この音だけでは動物にとって何の意味もない。しかし、訓練者は、この音を1次強化因子と関連づける。1次強化因子とは、餌や玩具等のように動物が本能的に報酬と感じるものである。クリッカーの音は、1次強化因子と多数回関連づけることで2次強化因子(条件性強化因子)となる。この第2強化因子は、動物にとって、もうすぐ報酬が貰えるというポジティブなヒントの役割をする。クリッカー自体は、動物にとって報酬ではなく、動物を所望の行動に導くために利用される。また、このクリッカーにより、どの行動が強化されるべきかを動物に知らせることができる。訓練者は、動物が所望の行動をした際にのみ1次強化因子を与える。これは、所望の行

動へと導く一連の誘導工程の終了を示す合図である。

【0087】クリッカートレーニングは、少なくとも以下に示す4つのステップで構成されている。

【0088】第1ステップ：クリッカーをチャージアップする。このステップにおいて、動物は、クリッカーの音をご褒美（餌等）と連鎖付ける必要がある。それには、クリッカーを鳴らし、続いて報酬を与えることを連続して多数回、例えば20～50回ほど、動物がクリッカーの音で明らかに興奮するようになるまで繰り返す。

【0089】第2ステップ：行動を捕える。次に、動物を所望の動作をとるよう誘導する。例えば、訓練者が動物に右回りに円を描いて回って欲しいと思った場合には、動物の頭が右に若干動いた時点でクリッカーを鳴らすことから始まる。動物が何度も頭を動かすようになったら、今度は身体を右に動かし始めた時だけにクリッカーを鳴らす。徐々にクリッカーを鳴らす基準を上げていき、完全に身体が1回転すると報酬を与える。

【0090】第3ステップ：コマンドワードを与える。動物が所望の行動を学習したらコマンドワードを1度だけ言う。コマンドワードを言うタイミングは、動物がその行動を行った直後または直前である。

【0091】第4ステップ：行動を試す。次に、学習した行動を試し、洗練する必要がある。動物が正確な行動をとった時にだけ、コマンドワードを言い、報酬を与える。

【0092】以上説明したように、クリッカートレーニングは、動物を所望の行動に導くためのものである。さらに、このクリッカートレーニングは、動物が自発的に行うことは滅多にない稀な行動を学習させるためや、連続した一連の行動を学習させるために使用することができる。

【0093】以下では、このクリッカートレーニングの原理を用いることで上述したロボット装置1が行動を学習する方法について説明する。

【0094】（4）ロボット装置への適用  
上述のロボット装置1において本発明が適用された要部を説明する。ロボット装置1は、画像信号や音声信号（音響信号）に関連づけて行動を学習し、学習後は、関連づけられた画像信号や音声信号により行動を引き起こすように構成されている。以下の説明では、学習として行動に音声に関連づける例について主に説明するが、行動に画像に関連づけることもできることはいうまでもない。具体的には、ロボット装置1は、本発明の実現のために次のような構成を有している。

【0095】ロボット装置1は、図9に示すように、入力部101と、1次強化因子検出部102と、2次強化因子連想部103と、行動決定部104と、行動記憶部105と、行動出力部106と、音声認識部107と、連想記憶部108とを備えている。

【0096】ここで、入力部101は、図2に示すマイ

クロホン23やタッチセンサ21である。入力部101は、音声信号又はセンサ信号を1次強化因子検出部102及び2次強化因子連想部103に供給する。また、入力部101は、音声信号を音声認識部107に供給する。

【0097】1次強化因子検出部102は、入力部101から供給された音声信号又はセンサ信号から1次強化因子を検出する。1次強化因子は、予めロボット装置1が報酬と感ずるよう設定されているものであり、例えば、ロボット装置1の頭部への圧力（頭を軽くたたくことによるもの）の検出や、おめでとうを表す音声信号等が挙げられる。以下では、1次強化因子として「ブラボー」という音声信号を用いる。1次強化因子検出部102は、1次強化因子を検出すると、1次強化信号を行動決定部104及び行動記憶部105に供給する。

【0098】2次強化因子連想部103は、入力部101から供給された音声信号又はセンサ信号から2次強化因子を連想して検出する。2次強化因子は、1次強化因子と同様に予めロボット装置1に設定されているものであってもよく、また、訓練者が1次強化因子に関連づけて設定するようにしてもよい。訓練者が2次強化因子を設定する場合には、例えば、ロボット装置1を2次強化因子設定モードとし、図10に示すようにして設定を行う。

【0099】図10に示すように、2次強化因子を設定する際には、先ずステップS1において、2次強化因子とする刺激を与える。この刺激は、接触刺激であっても音声刺激であっても構わない。

【0100】続いてステップS2において、1次強化因子を与え、ステップS3において、ステップS1で与えられた刺激を1次強化因子と関連づけて2次強化因子として記憶する。

【0101】なお、2次強化因子の設定手法は、この例に限定されず、例えば、ロボット装置1がある刺激の後に1次強化因子が与えられたことを多数回、例えば30回以上検出すると、その刺激を2次強化因子として記憶するように予め設定しておいてもよい。また、設定した2次強化因子を変更可能とすることもできる。この際、例えば、2次強化因子変更コマンドを与えながら任意の刺激を与えることにより、2次強化因子を変更することができる。

【0102】なお、この2次強化因子は、ロボット装置1を適当な行動へ誘導するための優れた標識として機能できるように、素早く与えることができ、且つ、検出し易いものでなくてはならない。以下では、2次強化因子として「グッド」という音声信号を用いる。

【0103】2次強化因子連想部103は、2次強化因子を検出すると、2次強化信号を行動決定部104及び行動記憶部105に供給する。

【0104】行動決定部104は、あるノード（状態）

から他のノードへの遷移確率に基づいて確率的に次の行動を決定し、決定結果を行動出力部 106 に供給する。また、行動決定部 104 は、決定された行動の履歴を行動記憶部 105 に供給して記憶させる。さらに、行動決定部 104 は、1 次強化因子検出部 102 及び 2 次強化因子連想部 103 から供給された 1 次強化信号及び 2 次強化信号に基づいて、後述するように、あるノード（状態）の選択確率が高くなるように遷移確率を更新するか、又は直前のノード（状態）からそのノード（状態）に遷移しやすくなるように遷移確率を更新する。

【0105】行動記憶部 105 は、行動決定部 104 において決定された行動の履歴を記憶する。また、行動記憶部 104 には、1 次強化因子検出部 102 及び 2 次強化因子連想部 103 からの 1 次強化信号及び 2 次強化信号が供給され、行動学習モードになってから 1 次強化信号が供給されるまでに 2 次強化信号が与えられた行動を一連の行動として記憶する。

【0106】行動出力部 106 は、行動決定部 105 から供給された行動の決定結果に基づいて、上述した行動コマンドを図 4 に示す出力セマンティクスコンバータモジュール 68 に送出する。

【0107】音声認識部 107 は、入力部 101 から入力された音声信号を音声処理して、所定の言語として認識する。具体的には、音声認識部 107 は、HMM (HiddenMarkov Model) を採用して構成されており、HMM による複数の認識クラスにより、入力された音声を音韻系列として認識する。

【0108】連想記憶部 108 は、行動学習モードになってから 1 次強化信号が供給されるまでに 2 次強化信号が与えられた一連の行動と音声認識部 107 で認識された当該行動を表すコマンドとを関連づけて連想記憶する。コマンドと関連づけて連想記憶された一連の行動は、行動記憶部 104 に記憶される。

【0109】ロボット装置 1 が以上のような構成を有することで、訓練者は、2 次強化因子を使ってロボット装置 1 を所望の行動へ誘導することができる。すなわち、訓練者は、ロボット装置 1 が所望の行動に近づくと 2 次強化因子を与え、その行動に近づくための標識とする。ロボット装置 1 が最終的に所望の行動を行うと、訓練者は、1 次強化因子を与え、ロボット装置は、それまでの 2 次強化因子の与えられた一連の行動を記憶する。そして、訓練者がコマンドを発声すると、ロボット装置は、記憶した一連の動作を順に出力する。なお、記憶された一連の行動を後述するようにして絞り込むことも可能である。

【0110】この行動学習方法の一連の手順を図 11 に示す。図 11 に示すように、先ずステップ S10 において、行動学習モードに設定される。

【0111】次にステップ S11 において、ノードからノードへの遷移確率が初期化され、ステップ S12 にお

いて、ロボット装置 1 は、現在の遷移確率に応じて行動を決定し、その行動を出力する。

【0112】ステップ S13 では、ステップ S12 で出力した行動によって 1 次強化因子が与えられたか否かが判別される。1 次強化因子が与えられた場合には、ステップ S17 に進む。ステップ S13 において、1 次強化因子が与えられていない場合には、ステップ S14 に進む。

【0113】ステップ S14 では、ステップ S12 で出力した行動によって 2 次強化因子が与えられたか否かが判別される。2 次強化因子が与えられていない場合には、所望の行動に近づくための行動ではないとし、ステップ S12 に戻って再び次の行動を決定する。ステップ S14 において、2 次強化因子が与えられた場合には、ステップ S15 に進む。

【0114】ステップ S15 では、ステップ S12 で出力した行動を所望の行動に至る行動として記憶し、ステップ S16 に進む。

【0115】ステップ S16 では、2 次強化信号が与えられた行動の選択確率が高くなるように遷移確率を更新してステップ S12 に戻る。

【0116】ステップ S17 では、1 次強化信号が与えられた信号と関連づけるためのコマンドを発声する。

【0117】ステップ S18 では、1 次強化信号が与えられた行動とその行動に至るまでに 2 次強化信号が与えられた行動とを一連の行動として、ステップ S17 で発声したコマンドと関連づけて記憶して終了する。

【0118】本実施の形態における行動学習方法では、以上のような手順により、ロボット装置 1 が所望の行動へ誘導され、その行動を学習する。

【0119】なお、図 11 のフローチャートでは、遷移確率を初期化した後に所望の行動と関連づけるコマンドを発声するようにしたが、これに限定されるものではなく、ロボット装置 1 が所望の行動を行った直後に、その行動と関連づけるコマンドを発声するようにしても構わない。

【0120】また、図 11 のフローチャートでは、ステップ S16 において、2 次強化信号が与えられた行動の選択確率が高くなるように遷移確率を更新するものとして説明したが、これに限定されるものではなく、例えば直前の行動から 2 次強化信号が与えられた行動に遷移しやすくなるように遷移確率を更新するようにしても構わない。

【0121】ところで、訓練者は、2 次強化因子を使ってロボット装置 1 を所望の動作へと誘導することができるが、これにはまず、ロボット装置 1 の行動の「位相」を定義しておく必要がある。すなわち、ロボット装置 1 は、ある行動をとった後にはそれと「類似した」行動への切替えが可能でなければならない。現存する行動の位相を定義するための方法として、以下に 2 つの例を挙げ

る。

【0122】先ず1番目の方法は、各行動を説明的特性と志向的特性とに分類される1組の特性を用いて説明するというものである。

【0123】説明的特性は、例えば開始姿勢（立つ、座る、横たわる）であり、実際、これにはロボットの頭、脚、またはその他の部分、あるいは音を発するか否かが関連している。

【0124】志向的特性は、行動を起こさせる目的、例えば、移動するためのものであるか、何かを掴むためのものであるか、あるいは注目を得るためのものであるかを表す。各行動は、これらの次元によって定義された空間中の点としてみる事ができる。この特性に従って全ての行動を形式化し、2つの行動間の「距離」を定義することで、ある行動から「類似した」行動へと誘導することができる。

【0125】この方法では、一旦特性が選択されると全ての行動をかなり短時間で表現できるが、行動間の遷移が予測し難い。

【0126】次に2番目の方法は、各行動間の遷移を示すグラフを作成するものである。この方法によれば、ロボット装置1がとる行動の遷移をより有効に把握することができる。第1番目の方法と同様に、行動間の客観的な類似点を、「志向性」と関連した何らかの基準と組み合わせることが可能である。さらに、一般的な行動（座る、立つ等）と稀な行動（特別なダンスを踊る、体操をする）とをより明確に区別することができる。図12に、グラフフォーマリズムで定義した位相の一部を示す。また、ロボット装置1が図12に挙げられた行動をとった状態の一部を図13に示す。

【0127】図12を用いて、訓練者が、ロボット装置1が減多にとらない特別な行動である穴掘り行動のための言葉を教える場合について具体的に説明する。この行動では、図13（G）に示すように、ロボット装置1は座っており、地面を引掻くのに左前脚を使用する。また、頭はその脚を見ており、その動きを追う。この行動は、図12のノード[DIG]の状態に対応する。

【0128】最初に、ロボット装置1は、図13（A）に示すように立ち上がっている、すなわちノード[STAND]の状態にあるとする。ロボット装置1は、先ず図13（B）に示すように歩行を始める、すなわち、ノード[STAND]の状態からノード[WALK]の状態に遷移する。このノード[WALK]への遷移は、所望の状態であるノード[DIG]に近づくものではないので、訓練者は、何の強化因子も与えない。そこでロボット装置1は、別の行動をノード[STAND]の状態から開始する。なお、この際、自動的にノード[STAND]の状態に戻るようにしてもよく、また、遷移確率に従って戻るようにしてもよい。

【0129】次にロボット装置1は、図13（D）に示

すように座る、すなわちノード[SIT]に遷移する。このノード[SIT]への遷移は、所望の状態であるノード[DIG]に近づくものであるため、訓練者は、「グッド」と言い、ロボット装置1に対して2次強化因子を与える。この2次強化因子が与えられることで遷移確率が更新され、ロボット装置1は、2次強化因子の与えられた行動である座る行動（ノード[SIT]）と関連づけられた行動をとろうとする。

【0130】続いてロボット装置1は、図13（E）に示すように、2本の前脚で押し始める。すなわちノード[PUSH]に遷移する。しかし、このノード[PUSH]への遷移は、所望の状態であるノード[DIG]に近づくものではないので、訓練者は何の強化因子も与えない。そこでロボット装置1は、別の行動をノード[SIT]の状態から開始する。この際にも、自動的にノード[SIT]の状態に戻るようにしてもよく、また、遷移確率に従って戻るようにしてもよい。

【0131】続いてロボット装置1は、ノード[HELLO]に遷移し、図13（F）に示すように左前脚を上げて「ハロー」の動作を行う。この行動は、穴掘りに使用する左前脚が関連しているため、訓練者は、ロボット装置1に「グッド」と言い、2次強化因子を与える。

【0132】以上のような訓練者とロボット装置1とのやりとりを経て、ロボット装置1がノード[DIG]に遷移して穴掘り行動を開始すると、訓練者は、「ブラボー」と言い、ロボット装置1に対して1次強化因子を与える。この1次強化因子は、ロボット装置1が所望の行動を行ったことを表す。以上の結果、ロボット装置1は、[SIT] - [HELLO] - [DIG]という一連の行動を記憶する。

【0133】なお、上述した[SIT] - [HELLO] - [DIG]という経路に限らず、別の経路を使っても、ロボット装置1を所望の行動へと誘導することができる。例えば、訓練者は、図13（C）に示すような蹴る行動、すなわちノード[KICK]を介してロボット装置1に左脚を移動させ、次に座るように誘導することができる。

【0134】所望の行動が実行され1次強化因子を与えると、訓練者は、即座に行動と関連づけるコマンドワードをロボット装置1に与える。その際、ロボット装置1がコマンドワードを確かに理解したことを確認するために、何らかのフィードバックを採用することが好ましい。フィードバックとしては、例えば、ロボット装置1が図2に示したスピーカ24を介してそのコマンドを真似して繰返し、確認を要求するようにしてもよく、また、「目」であるLEDを点滅させる等により、何かを理解した旨を示し、コマンドが再度繰り返されるのを待つようにしてもよい。ロボット装置1が、訓練者により再び与えられたコマンドを検出できない場合には、訓練者にもう一度同一のことをさせようとするための特別な

行動（例えば、頭を振る等）により、コマンドを理解できなかったことを表すようにしてもよい。

【0135】ロボット装置1は、このコマンドワードと上述した一連の行動とを関連づけて記憶し、再度コマンドワードが与えられた際には、記憶した一連の行動、例えば上述した[S I T] - [H E L L O] - [D I G]という一連の行動を行う。

【0136】ここで、訓練者が一連の行動のうちの穴掘り行動とのみコマンドワードを関連づけたい場合には、その一連の行動を絞り込む必要がある。このため、コマンドワードが与えられると、ロボット装置1は、記憶した一連の動作を自発的に繰り返すようにすることができる。ロボット装置1が一連の動作を再び実行した後に1次強化因子が与えられた場合には、ロボット装置1は、そのコマンドが一連の行動と関連していると判断する。そうでない場合には、ロボット装置1は、前に行った一連の行動に由来するがそれよりも段階数の少ない新しい一連の行動を作り出し、その新しい一連の行動を行う。このようにして、最終的に、ロボット装置1は、そのコマンドワードが穴掘り行動のみに関連したものであると判断することができる。

【0137】以上説明したように、本実施の形態におけるロボット装置1は、訓練者が動物や人に行動を教えるときと同様にして行動を教示することで、その行動を効率的に学習することができる。また、その際、2次強化因子を標識として所望の行動へ誘導されるため、複雑な動作や稀にしか行わない行動を学習するときの試行錯誤の回数を減らすことができる。さらに、2次強化因子が与えられた際に遷移確率を更新することによって、所望の行動に遷移する確率が高くなり、所望の行動により早く達することが可能となる。

【0138】このような機能がロボット装置1に組み込まれることで、ユーザ（飼主）とロボット装置1とが自然なコミュニケーションを図ることができる。なお、この制御プログラムは、図2のメモ리카ード28又はフラッシュROM12に格納される。

【0139】なお、本発明は上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。

【0140】例えば、上述した実施の形態では、行動と関連づけるコマンドとして音声情報を用いたが、これに限定されるものではなく、例えばタッチセンサを介した接触情報やCCDを介した画像情報であっても構わない。

【0141】また、上述した実施の形態では、クリックートレーニングをロボット装置に行動を教示する際に用いたが、これに限定されるものではなく、他の目的にも用いることができる。すなわち、物体の名前を教える場合に、ロボット装置が対象物に近づく際に2次強化因子を与えることで、ロボット装置は、その対象物に誘導さ

れる。これにより、共同注意といった困難な問題を軽減することができる。

【0142】また、クリックートレーニングと上述したルアーリングとを組み合わせるようにしても構わない。これにより、ある段階の行動が自発的に実行されるのを待つ必要がなくなる。

【0143】

【発明の効果】以上詳細に説明したように本発明に係るロボット装置は、動作部を制御して行動を出現させるロボット装置であって、行動の出現傾向に基づいて、一の行動を決定する行動決定手段と、上記行動決定手段が決定した行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力手段と、外部情報を入力する入力手段と、上記入力手段が入力した上記外部情報から予め設定された1次強化因子を検出する1次強化因子検出手段と、上記入力手段が入力した上記外部情報から2次強化因子を検出する2次強化因子検出手段と、上記行動決定手段が決定した行動を記憶する行動記憶手段とを備え、上記行動決定手段は、上記2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向を高め、上記行動記憶手段は、上記1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動を記憶することを特徴としている。

【0144】ここで、2次強化因子が検出された行動は、1次強化因子が検出された行動に至るまでの行動である。

【0145】このようなロボット装置は、行動の学習時に、2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向を高め、1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動を記憶する。

【0146】この発明が適用されたロボット装置は、2次強化因子を標識として学習しようとする行動に達することができる。行動を効率的に学習することができる。

【0147】また、上述した目的を達成するために、本発明に係るロボット装置の行動学習方法は、動作部を制御して行動を出現させるロボット装置の行動学習方法であって、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、外部情報を入力する入力手段が入力した上記外部情報から予め設定された1次強化因子を検出する1次強化因子検出工程と、上記入力手段が入力した上記外部情報から2次強化因子を検出する2次強化因子検出工程と、上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、上記行動決定工程では、上記2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、上記行動記憶工程では、上記1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記

憶されることを特徴としている。

【0148】ここで、2次強化因子が検出された行動は、1次強化因子が検出された行動に至るまでの行動である。

【0149】このようなロボット装置の行動学習方法では、行動の学習時に、2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記憶される。

【0150】この発明が適用されたロボット装置は、2次強化因子を標識として学習しようとする行動に達することができ、行動を効率的に学習することができる。

【0151】また、上述した目的を達成するために、本発明に係るロボット装置の行動学習プログラムは、動作部を制御して行動を出現させるロボット装置の行動学習プログラムであって、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、外部情報を入力する入力手段が入力した上記外部情報から予め設定された1次強化因子を検出する1次強化因子検出工程と、上記入力手段が入力した上記外部情報から2次強化因子を検出する2次強化因子検出工程と、上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、上記行動決定工程では、上記2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、上記行動記憶工程では、上記1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記憶されることを特徴としている。

【0152】ここで、2次強化因子が検出された行動は、1次強化因子が検出された行動に至るまでの行動である。

【0153】このようなロボット装置の行動学習プログラムでは、行動の学習時に、2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記憶される。

【0154】この発明が適用されたロボット装置は、2次強化因子を標識として学習しようとする行動に達することができ、行動を効率的に学習することができる。

【0155】また、上述した目的を達成するために、本発明に係るプログラム記録媒体は、動作部を制御して行動を出現させるロボット装置の行動学習プログラムが記録されたコンピュータ制御可能なプログラム記録媒体であって、上記ロボット装置の行動学習プログラムは、行動の出現傾向に基づいて、一の行動を決定する行動決定工程と、上記行動決定工程で決定された行動に基づいて、上記動作部を制御して、当該決定された行動を出現させる行動出力工程と、外部情報を入力する入力手段が入力した上記外部情報から予め設定された1次強化因子

を検出する1次強化因子検出工程と、上記入力手段が入力した上記外部情報から2次強化因子を検出する2次強化因子検出工程と、上記行動決定工程で決定された行動を記憶する行動記憶工程とを有し、上記行動決定工程では、上記2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、上記行動記憶工程では、上記1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記憶されることを特徴とするロボット装置の行動学習プログラムが記録されたものである。

【0156】ここで、2次強化因子が検出された行動は、1次強化因子が検出された行動に至るまでの行動である。

【0157】このようなプログラム記録媒体に記録されているロボット装置の行動学習プログラムでは、行動の学習時に、2次強化因子が検出される毎に、当該2次強化因子が検出された行動の出現傾向が高められ、1次強化因子が検出されると、少なくとも当該1次強化因子が検出された行動が記憶される。

【0158】この発明が適用されたロボット装置は、2次強化因子を標識として学習しようとする行動に達することができ、行動を効率的に学習することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態におけるロボット装置の外観構成を示す斜視図である。

【図2】同ロボット装置の回路構成を示すブロック図である。

【図3】同ロボット装置のソフトウェア構成を示すブロック図である。

【図4】同ロボット装置のソフトウェア構成におけるミドル・ウェア・レイヤの構成を示すブロック図である。

【図5】同ロボット装置のソフトウェア構成におけるアプリケーション・レイヤの構成を示すブロック図である。

【図6】同アプリケーション・レイヤの行動モデルライブラリの構成を示すブロック図である。

【図7】同ロボット装置の行動決定のための情報となる有限確率オートマトンを説明する図である。

【図8】有限確率オートマトンの各ノードに用意された状態遷移表を示す図である。

【図9】同ロボット装置における本発明に係る部分の構成を示すブロック図である。

【図10】2次強化因子を設定する場合の手順を説明するフローチャートである。

【図11】同ロボット装置の行動学習方法の一連の手順を説明するフローチャートである。

【図12】同ロボット装置が行う行動間の遷移を説明する図である。

【図13】同ロボット装置が行う行動の一部を説明する図である。

29

30

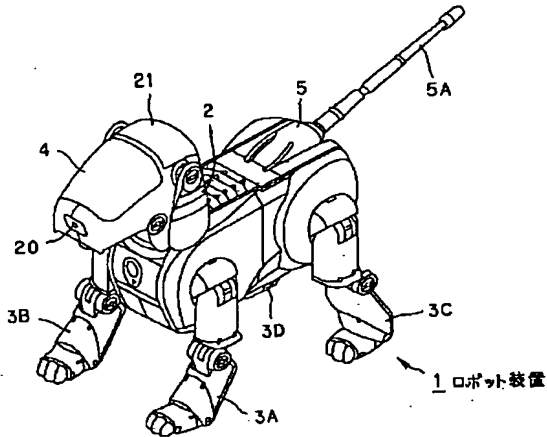
【図14】従来の動物に対する行動指示方法の例を説明する図である。

【符号の説明】

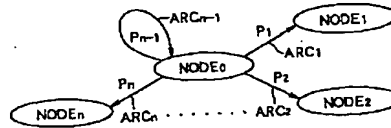
1 ロボット装置、101 入力部、102 1次強化

因子検出部、103 2次強化因子連想部、104 行動決定部、105 行動記憶部、106 行動出力部、107 音声認識部、108 連想記憶部

【図1】



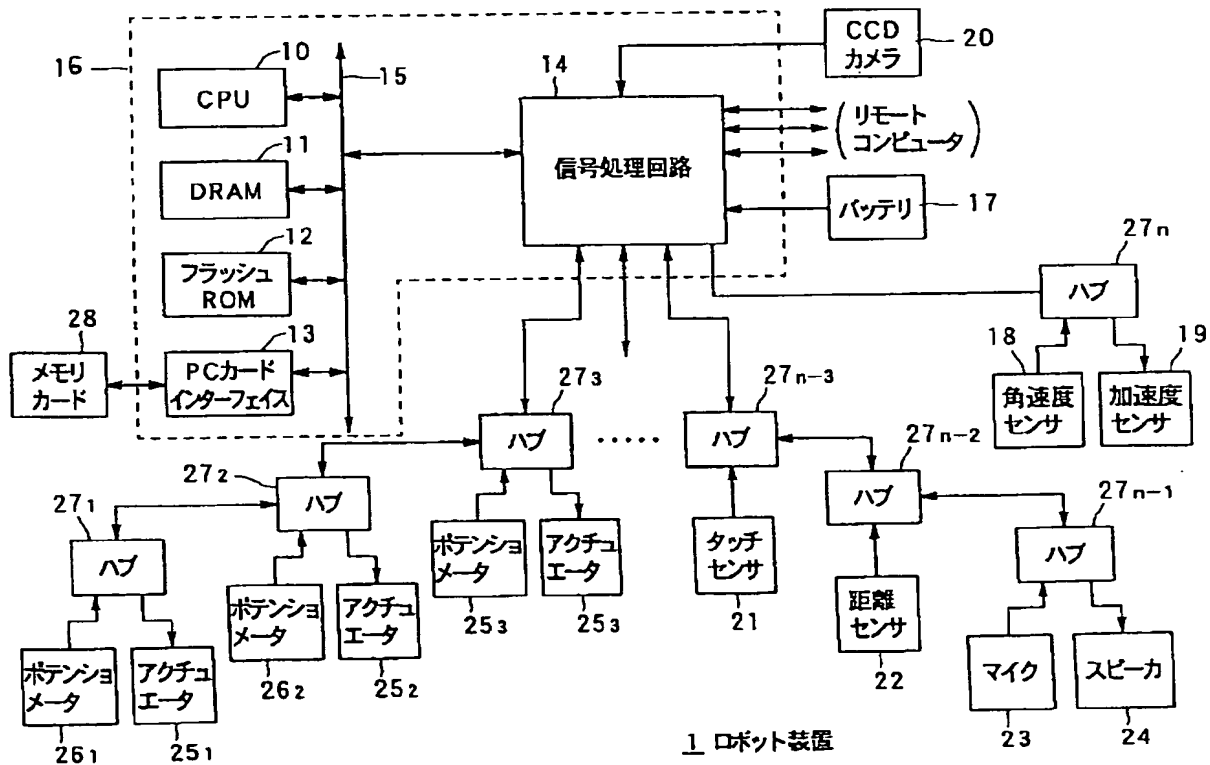
【図7】



【図14】

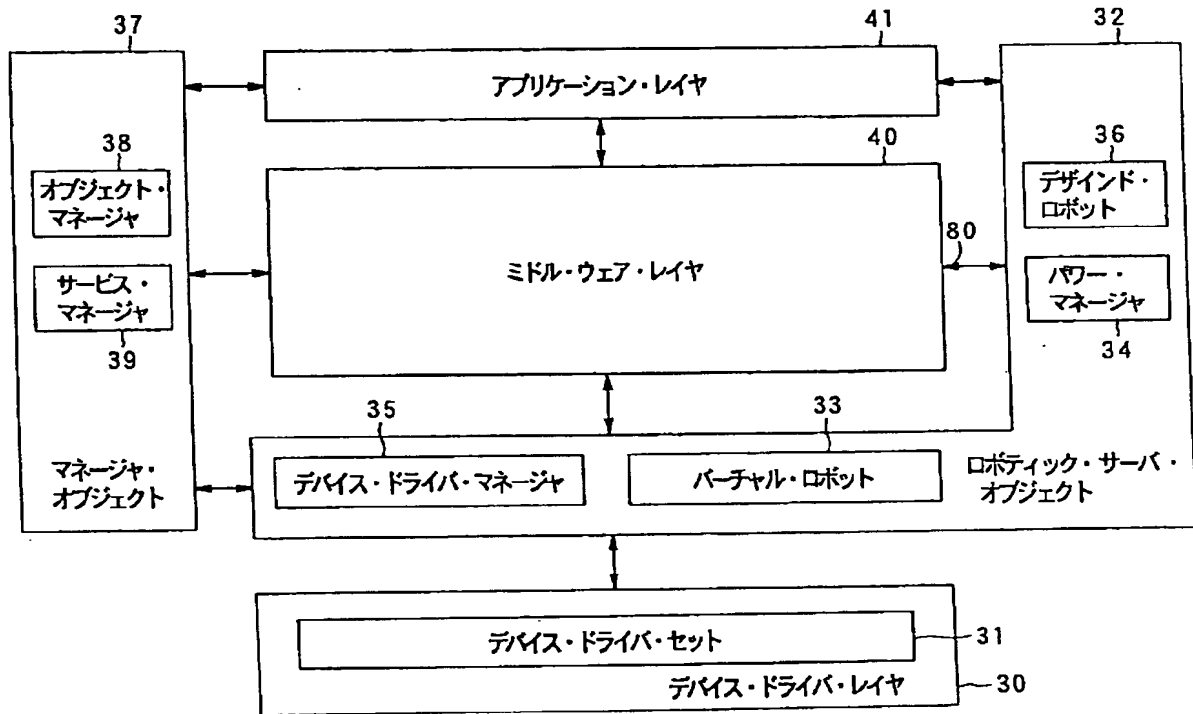
トレーニング技術	動作の連続性	珍しい動作	動物への適用
モデリング	なし	困難	希に使用される
ルアーリング	困難	困難	単純な動作はOK
キャプチャリング	なし	なし	良い
イミテーション	あり	あり	希に使用される

【図2】

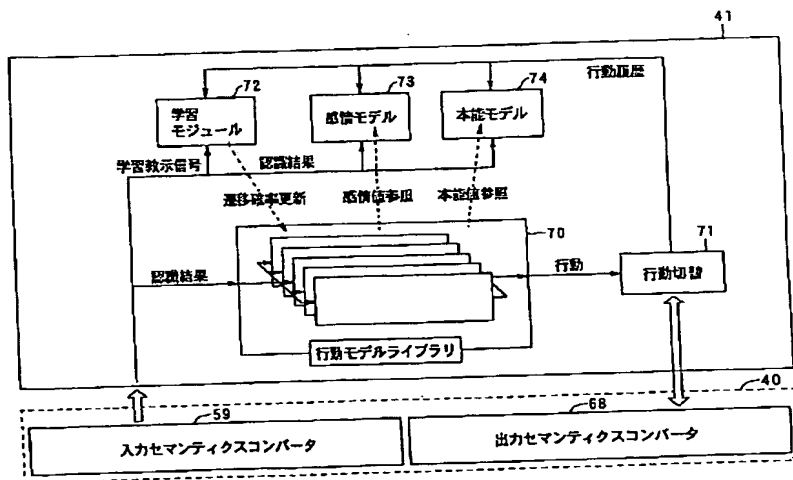




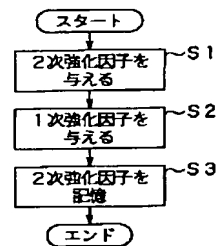
【図3】



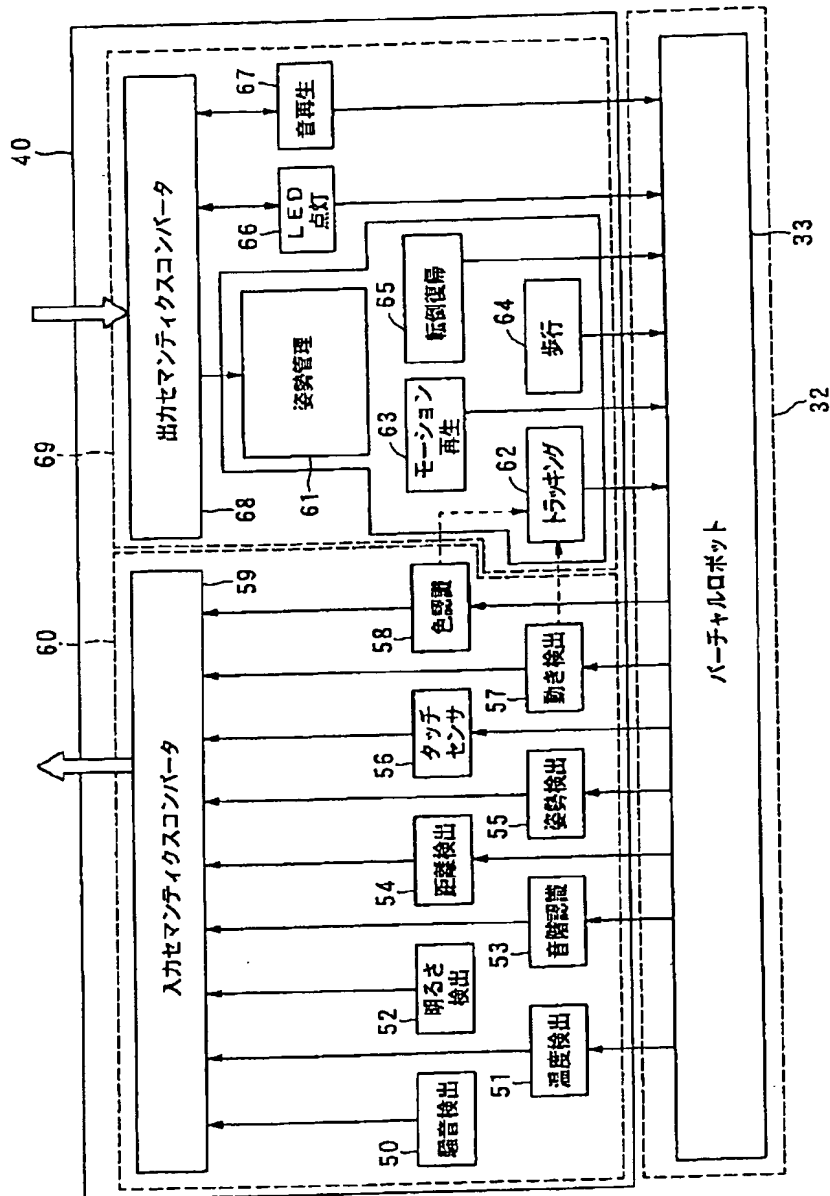
【図5】



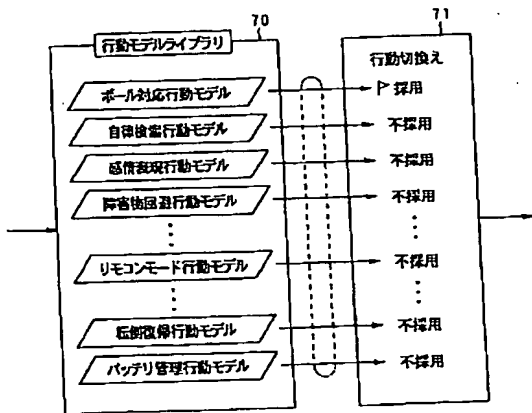
【図10】



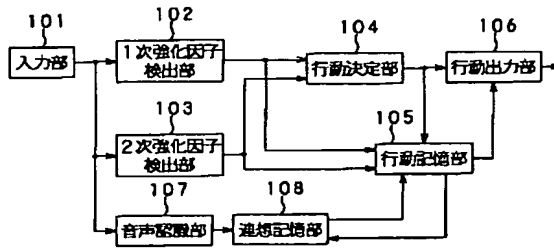
【図4】



【図6】



【図9】

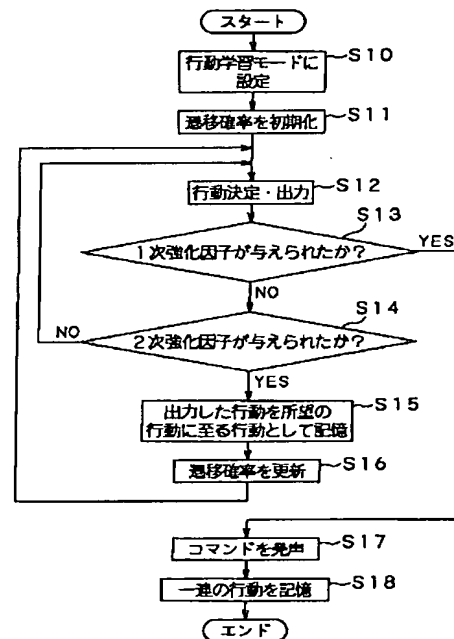


【図8】

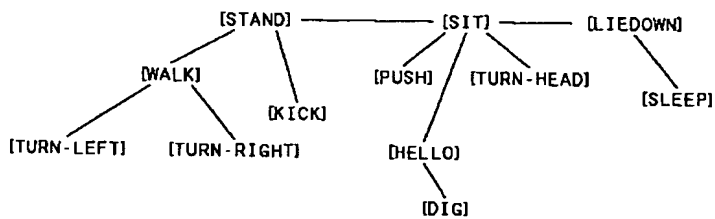
node 100	入力イベント名	データ名	データの範囲	他のノードへの遷移確率				DI	n
				A	B	C	D		
遷移先ノード				node 120	node 120	node 1000			node 500
出力行動				ACTION 1	ACTION 2	MOVE BACK			ACTION 4
1	BALL	SIZE	0.1000	30%					
2	PAT				40%				
3	HIT				20%				
4	MOTION					50%			
5	OBSTACLE	DISTANCE	0.100			100%			
6		JOY	50.100						
7		SURPRISE	50.100						
8		SADNESS	50.100						

80

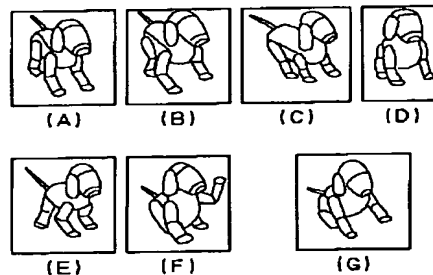
【図11】



【図12】



【図13】



## フロントページの続き

(72)発明者 カブラン フレデリック  
フランス国、75005 パリ市 アミヨ通り  
6番地 ソニーコンピュータサイエンス  
研究所 パリオフィス内  
(72)発明者 ピエール イブス オードエ  
フランス国、75005 パリ市 アミヨ通り  
6番地 ソニーコンピュータサイエンス  
研究所 パリオフィス内

(72)発明者 花形 理  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

Fターム(参考) 2C150 CA02 DA05 DA24 DA25 DA26  
DA27 DA28 DF03 DF04 DF06  
DF33 ED10 ED39 ED42 ED47  
ED52 EF07 EF09 EF16 EF17  
EF22 EF23 EF28 EF29 EF33  
EF36  
3C007 AS36 CS08 LW12 MT14 WA04  
WA14 WB18

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-039363

(43)Date of publication of application : 13.02.2003

---

(51)Int.Cl. B25J 13/00

A63H 11/00

B25J 5/00

---

(21)Application number : 2002-127374 (71)Applicant : SONY FRANCE SA  
SONY CORP

(22)Date of filing : 26.04.2002 (72)Inventor : KAPLAN FREDERIC

OUDEYER PIERRE-YVES

HANAGATA OSAMU

---

(30)Priority

Priority number : 2001 01401127

Priority date : 30.04.2001

Priority country : EP

---

(54) ROBOT DEVICE, ACTION LEARNING METHOD THEREFOR, ACTION  
LEARNING PROGRAM THEREOF, AND PROGRAM RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To efficiently learn complicated and continuous actions.

SOLUTION: In a robot device, a command for executing the action determined stochastically in an action determining part 104 is outputted by an action output

part 106 based on the action. When approaching desired action, a trainee gives secondary reinforcing factor to the robot device to use it as a mark for approaching the desired action. When the secondary reinforcing factor is detected, the action determining part 104 changes transition probability to the desired action. When the robot device performs the desired action finally, the trainee gives primary reinforcing factor to the robot device. When a series of actions reaching the desired action are related to commands, a series of actions read from an action storage part 105 and a command recognized in a voice recognition part 107 are associated and stored in an associative storage part 108.

---

LEGAL STATUS [Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

**\* NOTICES \***

JPO and NCIPPI are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.

2. \*\*\*\* shows the word which can not be translated.

3. In the drawings, any words are not translated.

---

**CLAIMS**

---

[Claim(s)]

[Claim 1] An action decision means to be robot equipment which a right hand side is controlled [ robot ] and makes action appear, and to opt for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which the above-mentioned action decision means opted. An action output means to make the action concerned for



which it opted appear, and an input means to input external information, A primary reinforcer detection means to detect the primary reinforcer beforehand set up from the above-mentioned external information which the above-mentioned input means inputted, A secondary reinforcement factor detection means to detect a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has an action storage means to memorize the action for which the above-mentioned action decision means opted. The above-mentioned action decision means It is robot equipment with which the appearance inclination of action for the secondary reinforcement factor concerned to have been detected whenever the above-mentioned secondary reinforcement factor was detected is raised, and the above-mentioned action storage means is characterized by memorizing the action by which the primary reinforcer concerned was detected at least if the above-mentioned primary reinforcer is detected.

[Claim 2] The action by which the above-mentioned secondary reinforcement factor was detected is robot equipment according to claim 1 characterized by being action until it results in the action by which the above-mentioned primary reinforcer was detected.

[Claim 3] The above-mentioned action storage means is robot equipment according to claim 1 characterized by memorizing the action by which the

above-mentioned primary reinforcer was detected on the basis of relevance with the action by which the above-mentioned secondary reinforcement factor was detected before the action concerned.

[Claim 4] The above-mentioned action storage means is robot equipment according to claim 3 characterized by memorizing the action by which the above-mentioned secondary reinforcement factor was detected by the action by which the above-mentioned primary reinforcer was detected as sequence action.

[Claim 5] The above-mentioned action decision means is robot equipment according to claim 4 characterized by raising the appearance inclination of each action included in the above-mentioned sequence action.

[Claim 6] If the above-mentioned primary reinforcer is detected, it will be made as [ reproduce / the above-mentioned sequence action / in order ]. The above-mentioned action output means When each action included in the above-mentioned sequence action is made to appear in order, consequently the above-mentioned primary reinforcer is detected When the above-mentioned sequence action is decided and memorized and the above-mentioned primary reinforcer is not detected, the above-mentioned action storage means The above-mentioned action output means is robot equipment according to claim 4 characterized by reducing the number of each actions included in the above-mentioned sequence action until the above-mentioned primary reinforcer

is detected, and making it appear in order.

[Claim 7] Robot equipment according to claim 1 characterized by having the associative storage means which associates and carries out associative storage of the action by which the above-mentioned primary reinforcer which the predetermined external information and the above-mentioned action storage means which the above-mentioned input means inputted memorized was detected.

[Claim 8] It is robot equipment according to claim 7 which is equipped with a speech-recognition means carries out speech processing of the speech information which the above-mentioned input means inputted, and recognize as predetermined language, and is characterized by for the above-mentioned associative-storage means to associate and carry out the associative storage of the action by which the above-mentioned primary reinforcer which the predetermined command and the above-mentioned action storage means recognized by the above-mentioned speech processing memorized was detected.

[Claim 9] The above-mentioned predetermined command is robot equipment according to claim 8 characterized by being inputted after the above-mentioned primary reinforcer is detected.

[Claim 10] The robot equipment according to claim 4 characterized by to have

the associative-storage means which associates and carries out the associative storage of the above-mentioned sequence action which a speech-recognition means carries out speech processing of the speech information which the above-mentioned input means inputted, and recognize as predetermined language, and the predetermined command and the above-mentioned action storage means in connection with the above-mentioned sequence action recognized by the above-mentioned speech processing memorized.

[Claim 11] It is robot equipment according to claim 8 with which it has the action learning mode for learning action, and the above-mentioned action storage means is characterized by memorizing the action by which the primary reinforcer concerned was detected at least if the above-mentioned primary reinforcer is detected in the above-mentioned action learning mode.

[Claim 12] The above-mentioned predetermined command is robot equipment according to claim 11 characterized by being inputted immediately after considering as the above-mentioned action learning mode.

[Claim 13] Robot equipment according to claim 1 which will be characterized by using the external information concerned as the above-mentioned secondary reinforcement factor if the above-mentioned primary reinforcer is detected after having the secondary reinforcement factor setting mode for setting up a secondary reinforcement factor and inputting the external information of

arbitration in the above-mentioned secondary reinforcement factor setting mode.

[Claim 14] Robot equipment according to claim 1 which will be characterized by using the external information concerned as the above-mentioned secondary reinforcement factor if that the above-mentioned primary reinforcer was detected continues more than the count of predetermined after the external information of arbitration is inputted.

[Claim 15] The above-mentioned primary reinforcer and the above-mentioned secondary reinforcement factor are robot equipment according to claim 1 characterized by being the speech information which the above-mentioned input means inputted.

[Claim 16] The action decision process of being the action study approach of the robot equipment which a right hand side is controlled [ robot ] and makes action appearing, and opting for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the

above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process The action study approach of the robot equipment characterized by memorizing the action by which the primary reinforcer concerned was detected at least if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised and the above-mentioned primary reinforcer is detected at the above-mentioned action storage process whenever the above-mentioned secondary reinforcement factor is detected.

[Claim 17] The action by which the above-mentioned secondary reinforcement factor was detected is the action study approach of the robot equipment according to claim 16 characterized by being action until it results in the action by which the above-mentioned primary reinforcer was detected.

[Claim 18] The action study approach of robot equipment according to claim 17 that action by which the above-mentioned primary reinforcer was detected is characterized by what is memorized by the basis of relevance with the action by which the above-mentioned secondary reinforcement factor was detected before the action concerned at the above-mentioned action storage process.

[Claim 19] The action study approach of the robot equipment according to claim

18 characterized by memorizing the action by which the above-mentioned secondary reinforcement factor was detected at the above-mentioned action storage process by the action by which the above-mentioned primary reinforcer was detected as sequence action.

[Claim 20] The action study approach of the robot equipment according to claim 19 characterized by raising the appearance inclination of each action included in the above-mentioned sequence action at the above-mentioned action decision process.

[Claim 21] If the above-mentioned primary reinforcer is detected, it will be made as [ reproduce / the above-mentioned sequence action / in order ]. At the above-mentioned action output process When each action included in the above-mentioned sequence action appears in order, consequently the above-mentioned primary reinforcer is detected When the above-mentioned sequence action is decided and memorized at the above-mentioned action storage process and the above-mentioned primary reinforcer is not detected The action study approach of the robot equipment according to claim 19 characterized by reducing the number of each actions included in the above-mentioned sequence action at the above-mentioned action output process until the above-mentioned primary reinforcer is detected, and appearing in order.

[Claim 22] The action study approach of the robot equipment according to claim 16 characterized by having the associative storage process which associates and carries out associative storage of the action by which the above-mentioned primary reinforcer memorized at the predetermined external information and the above-mentioned action storage process that the above-mentioned input means inputted was detected.

[Claim 23] The action study approach of the robot equipment according to claim 22 which has the speech-recognition process which carries out the speech processing of the speech information which the above-mentioned input means inputted, and recognizes as predetermined language, and is characterized at the above-mentioned associative-storage process by to associate and carry out the associative storage of the action by which the above-mentioned primary reinforcer memorized at the predetermined command and the above-mentioned action storage process recognized by the above-mentioned speech processing was detected.

[Claim 24] The above-mentioned predetermined command is the action study approach of the robot equipment according to claim 23 characterized by being inputted after the above-mentioned primary reinforcer is detected.

[Claim 25] The action study approach of the robot equipment according to claim 19 characterized by to have the associative-storage process which associates



and carries out the associative storage of the above-mentioned sequence action memorized at the speech-recognition process which carries out speech processing of the speech information which the above-mentioned input means inputted, and recognizes as predetermined language, and the predetermined command and the above-mentioned action storage process in connection with the above-mentioned sequence action recognized by the above-mentioned speech processing.

[Claim 26] The action study approach of the robot equipment according to claim 23 characterized by memorizing the action by which the primary reinforcer concerned was detected at least if it has the action learning mode for learning action and the above-mentioned primary reinforcer is detected in the above-mentioned action learning mode at the above-mentioned action storage process.

[Claim 27] The above-mentioned predetermined command is the action study approach of the robot equipment according to claim 26 characterized by being inputted immediately after considering as the above-mentioned action learning mode.

[Claim 28] The action study approach of the robot equipment according to claim 16 which will be characterized by using the external information concerned as the above-mentioned secondary reinforcement factor if the above-mentioned

primary reinforcer is detected after having the secondary reinforcement factor setting mode for setting up a secondary reinforcement factor and inputting the external information of arbitration in the above-mentioned secondary reinforcement factor setting mode.

[Claim 29] The action study approach of the robot equipment according to claim 16 which will be characterized by using the external information concerned as the above-mentioned secondary reinforcement factor if that the above-mentioned primary reinforcer was detected continues more than the count of predetermined after the external information of arbitration is inputted.

[Claim 30] The above-mentioned primary reinforcer and the above-mentioned secondary reinforcement factor are the action study approach of the robot equipment according to claim 16 characterized by being the speech information which the above-mentioned input means inputted.

[Claim 31] The action decision process of being the action learning program of robot equipment which a right hand side is controlled [ learning program ] and makes action appearing, and opting for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an

input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process Whenever the above-mentioned secondary reinforcement factor is detected, the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised. At the above-mentioned action storage process The action learning program of the robot equipment characterized by memorizing the action by which the primary reinforcer concerned was detected at least if the above-mentioned primary reinforcer is detected.

[Claim 32] It is the program documentation medium by which the action learning program of robot equipment which a right hand side is controlled [ learning program ] and makes action appear was recorded and in which computer control is possible. The action decision process that the action learning program of the above-mentioned robot equipment opts for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process.

The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process Whenever the above-mentioned secondary reinforcement factor is detected, the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised. At the above-mentioned action storage process The program documentation medium by which the action learning program of the robot equipment characterized by memorizing the action by which the primary reinforcer concerned was detected at least if the above-mentioned primary reinforcer is detected was recorded.

---

DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the program documentation medium by which the action study approach of the robot equipment which acts autonomously, and the robot equipment with which such robot equipment learns desired action, the action learning program of robot equipment, and the action learning program of the robot equipment concerned were recorded.

[0002]

[Description of the Prior Art] Conventionally, breaking is performed in order to tame animals, such as a dog. For example, the owner of a dog teaches action of a "hand", "stability", etc., etc. to a dog. However, the owner of many dogs has made the mistake of ordering coincidence, making it attach a dog to a desired location. For example, a trainer depresses the hips of a dog toward the ground, repeating the word "stability." It is from the following reasons that this approach is not successful.

[0003] (1) An animal has forced the selection which pays a trainer's words or either of the learned behavior attention.

[0004] (2) In order to repeat an instruction two or more times, an animal does not understand the part of a throat for whether it is connected with action before long.

[0005] (3) Before acting, it is ordered in very many cases. For example, since

"stability" is ordered while the animal still stands, "stability" cannot be related with the condition of having sat down.

[0006] For such a reason, many trainers teach an instruction and action separately. Action is taught first in fact and an instruction is taught below.

[0007] There are modeling (Modelling), a lure ring (Ruring), a capture ring (Capturing), an imitation (Imitating), etc. in the approach for teaching action to an animal as a typical thing, and it has the description as shown in drawing 14 , respectively. Hereafter, it explains briefly.

[0008] Although modeling is an approach which the owner of many dogs tries, that the trainer who is an expert carries out never does not have it. This approach will praise, if an animal can be physically manipulated to a desired position and that position can be reached. By this approach, it is an animal in the passive state. Therefore, in many cases, the efforts of action study do not bear fruit. Moreover, it is not easy to realize the complicated motion of the more than attained to a quiescence position by this approach.

[0009] RUARINGU is similar with modeling except for a point without physical contact with an animal. A trainer can place a toy and food before the nose of an animal, and can guide an animal to a desired position using this. Although a result with this sufficient approach in the case of an actual animal is obtained, that use is limited to the time when a certain position and very simple actuation

are taught.

[0010] Unlike the modeling and RUARINGU which were mentioned above, a capture ring uses spontaneous action of an animal. For example, whenever a trainer recognizes it as the animal taking lessons from the desired position, or having taken exact action, he gives positive reinforcer to an animal. Although the result in which it is good also about this approach in the case of an actual animal is obtained, he needs to wait until an animal attaches a trainer to a desired position. For example, when teaching "stability", a trainer needs to wait for an animal to sit down spontaneously. Moreover, this approach is difficult to teach continuous actuation and new actuation in order to use spontaneous action of an animal.

[0011] Although an imitation makes an animal copy the same actuation as a trainer, that a trainer uses the approach based on an imitation does not have it rash. Although it can teach also about continuous actuation or new actuation by this approach, this approach is checked only about higher animals, such as the primates and a marine mammal.

[0012]

[Problem(s) to be Solved by the Invention] By the way, the robot equipment made into the configuration which imitated the animal, and the so-called pet robot are proposed in recent years. Such robot equipment has a configuration

similar to the dog bred at ordinary homes, or a cat, and acts autonomously according to "it strikes" and the influence of "stroking" from a user (owner), a surrounding environment, etc. For example, action of "barking", "sleeping", etc. is carried out like an actual animal as autonomous action.

[0013] If such robot equipment can approach action like an actual animal more, the feeling of a living thing of robot equipment will increase further, and a user (owner) will sense the much more sense of closeness and satisfaction for robot equipment. Thereby, the amusement nature of robot equipment improves.

[0014] For example, if a user (owner) can teach desired actuation to the robot equipment which oneself owns and "can accustom" robot equipment the same with receiving an actual animal, a user (owner) will be considered to come to sense the much more sense of closeness and satisfaction for robot equipment.

[0015] However, in order to apply the approach for teaching action to an actual animal which was mentioned above to the robot equipment which acts autonomously, various problems exist.

[0016] For example, many modeling mentioned above has been used in order to teach a position to robot equipment in the industrial world, but if robot equipment acts autonomously and always becomes active, a problem will produce it. That is, robot equipment can make a judgment which will sit down if it can sense that the trainer is pushing the back and is programmed. However, there is a problem that



it is not easy to realize the complicated motion of the more than attained to a quiescence position by this approach.

[0017] Moreover, RUARINGU mentioned above is seldom used by robotics. If it is programmed so that robot equipment shows interest automatically to a predetermined thing, for example, a red thing, a trainer can make robot equipment attach to a desired position using this property. However, this also has the problem that the use range is restricted.

[0018] The capture ring mentioned above can create the model of sending that signal later, whenever robot equipment takes action of the request relevant to a signal (for example, language) using this approach. However, in order to teach language like "stability" to robot equipment, a trainer has the problem that it is necessary to wait for robot equipment to sit down spontaneously. Moreover, there are a problem that there is too many actuation which can name, and a problem that it is difficult to teach continuous actuation and new actuation in order to use spontaneous action of an animal.

[0019] the imitation mentioned above -- for example, -- "-- Y. -- in robotics, some research consortia have tackled as indicated by Kuniyoshi et al., "Learning by watching:Extracting reusable task knowledge from visual observation of human performance.", IEEE Transactions on Robotics and Automation, 10(6):799-822, 1994", etc. Although the continuous actuation and very new actuation can be

theoretically taught by this approach, since a great quantity of count capacity is needed, it is difficult to apply this approach to the existing robot equipment which acts autonomously.

[0020] This invention is proposed in view of such the conventional actual condition, and aims at offering the program documentation medium by which the action study approach of the robot equipment which can learn complicated and continuous action etc., and such robot equipment, the action learning program of robot equipment, and the action learning program of the robot equipment concerned were recorded.

[0021]

[Means for Solving the Problem] In order to attain the purpose mentioned above, the robot equipment concerning this invention An action decision means to be robot equipment which a right hand side is controlled [ robot ] and makes action appear, and to opt for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which the above-mentioned action decision means opted. An action output means to make the action concerned for which it opted appear, and an input means to input external information, A primary reinforcer detection means to detect the primary reinforcer beforehand set up from the above-mentioned external information which the above-mentioned input means inputted, A secondary

reinforcement factor detection means to detect a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has an action storage means to memorize the action for which the above-mentioned action decision means opted. The above-mentioned action decision means Whenever the above-mentioned secondary reinforcement factor is detected, the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised, and the above-mentioned action storage means is characterized by the primary reinforcer concerned memorizing the action by which it was detected at least, if the above-mentioned primary reinforcer is detected.

[0022] Here, the action by which the secondary reinforcement factor was detected is action until it results in the action by which primary reinforcer was detected.

[0023] Such robot equipment will memorize the action by which the primary reinforcer concerned was detected at least, if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it, whenever a secondary reinforcement factor is detected.

[0024] In order to attain the purpose mentioned above, moreover, the action study approach of the robot equipment concerning this invention The action

decision process of being the action study approach of the robot equipment which a right hand side is controlled [ robot ] and makes action appearing, and opting for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process If the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised and the above-mentioned primary reinforcer is detected at the above-mentioned action storage process whenever the above-mentioned secondary reinforcement factor is detected, it is characterized by memorizing the action by which the primary reinforcer concerned was detected at least.

[0025] Here, the action by which the secondary reinforcement factor was

detected is action until it results in the action by which primary reinforcer was detected.

[0026] By such action study approach of robot equipment, if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it whenever a secondary reinforcement factor is detected, the action by which the primary reinforcer concerned was detected at least will be memorized.

[0027] In order to attain the purpose mentioned above, moreover, the action learning program of the robot equipment concerning this invention The action decision process of being the action learning program of robot equipment which a right hand side is controlled [ learning program ] and makes action appearing, and opting for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the

above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process If the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised and the above-mentioned primary reinforcer is detected at the above-mentioned action storage process whenever the above-mentioned secondary reinforcement factor is detected, it is characterized by memorizing the action by which the primary reinforcer concerned was detected at least.

[0028] Here, the action by which the secondary reinforcement factor was detected is action until it results in the action by which primary reinforcer was detected.

[0029] In such an action learning program of robot equipment, if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it whenever a secondary reinforcement factor is detected, the action by which the primary reinforcer concerned was detected at least will be memorized.

[0030] In order to attain the purpose mentioned above, moreover, the program documentation medium concerning this invention It is the program

documentation medium by which the action learning program of robot equipment which a right hand side is controlled [ learning program ] and makes action appear was recorded and in which computer control is possible. The action decision process that the action learning program of the above-mentioned robot equipment opts for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process Whenever the above-mentioned secondary reinforcement factor is detected, the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised. At the above-mentioned action storage process Detection of the above-mentioned primary reinforcer records the action learning program of the robot equipment

characterized by memorizing the action by which the primary reinforcer concerned was detected at least.

[0031] Here, the action by which the secondary reinforcement factor was detected is action until it results in the action by which primary reinforcer was detected.

[0032] In the action learning program of the robot equipment currently recorded on such a program documentation medium, if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it whenever a secondary reinforcement factor is detected, the action by which the primary reinforcer concerned was detected at least will be memorized.

[0033]

[Embodiment of the Invention] Hereafter, the gestalt of the concrete operation which applied this invention is explained to a detail, referring to a drawing. The gestalt of this operation applies this invention to the robot equipment which acts autonomously according to a surrounding environment (external factor) or an internal condition (inner factor).

[0034] First, the gestalt of operation explains the configuration of robot equipment and explains the application part of this invention in robot equipment to a detail after that.



[0035] (1) As shown in the block diagram 1 of the robot equipment in the gestalt of this operation, while considering as the so-called pet robot of the configuration which imitated the "dog" and connecting the leg units 3A, 3B, and 3C and 3D with front and rear, right and left of the idiosoma unit 2, respectively, the head unit 4 and the tail section unit 5 are connected and constituted by the front end section and the back end section of the idiosoma unit 2, respectively.

[0036] As shown in drawing 2 , the control section 16 formed by connecting CPU (Central Processing Unit)10, DRAM (Dynamic Random Access Memory)11, a flash ROM (Read Only Memory) 12, PC (Personal Computer) card interface circuitry 13, and a digital disposal circuit 14 mutually through an internal bus 15 and the dc-battery 17 as a source of power of this robot equipment are contained by the idiosoma unit 2. Moreover, the angular-velocity sensor 18, an acceleration sensor 19, etc. for detecting the sense of robot equipment 1 and the acceleration of a motion are contained by the idiosoma unit 2.

[0037] Moreover, the CCD (Charge Coupled Device) camera 20 for picturizing an external situation to the head unit 4, The touch sensor 21 for detecting the pressure received by "it strokes" and the physical influence of "striking" from a user, The distance robot 22 for measuring the distance to the body located ahead, LED (Light Emitting Diode) (not shown) equivalent to the microphone 23 for collecting alien frequencies, the loudspeaker 24 for outputting voice, such as

a cry, and the "eye" of robot equipment 1 etc. is arranged in the predetermined location, respectively.

[0038] Furthermore, the actuators 251 and 252 for free frequency, ... and potentiometers 261 and 262, and ... are arranged in the joining segment of the joint part of each leg unit 3A-3D, each joining segment of each leg unit and the idiosoma unit 2, the head unit 4, and the idiosoma unit 2, and the list by the joining segment of tail 5A of the tail section unit 5, respectively. For example, actuators 251 and 252 and ... have the servo motor as a configuration. Leg unit 3A - 3D are controlled by the drive of a servo motor, and it changes in the target posture or actuation.

[0039] And it connects with the digital disposal circuit 14 of the control section 16 through the hubs 271-27n corresponding to various sensor lists, such as these angular-velocity sensor 18, an acceleration sensor 19, a touch sensor 21, a distance robot 22, a microphone 23, a loudspeaker 24 and each potentiometers 261 and 262, and ..., in LED and each actuators 251 and 252, and ..., respectively, and direct continuation of CCD camera 20 and the dc-battery 17 is carried out to the digital disposal circuit 14, respectively.

[0040] A digital disposal circuit 14 incorporates sensor data, and the image data and voice data which are supplied from each above-mentioned sensor one by one, and carries out sequential storing of these through an internal bus 15 in the

predetermined location in DRAM11, respectively. Moreover, a digital disposal circuit 14 incorporates the dc-battery residue data showing the dc-battery residue supplied from a dc-battery 17 with this one by one, and stores this in the predetermined location in DRAM11.

[0041] Thus, each sensor data stored in DRAM11, image data, voice data, and dc-battery residue data are used in case CPU10 performs motion control of this robot equipment 1 after this.

[0042] CPU10 stores direct read-out and this in DRAM11 for the control program stored in the memory card 28 or flash ROM 12 with which the PC Card slot which the idiosoma unit 2 does not illustrate was loaded through the PC card interface circuitry 13 at the time of the first stage when the power source of robot equipment 1 was switched on in practice.

[0043] Moreover, as mentioned above, from a digital disposal circuit 14, each sensor data by which sequential storing is carried out, image data, voice data, and dc-battery residue data are based on DRAM11, and CPU10 judges the situation of self and a perimeter, the existence of the command from a user, and influence, etc. after this.

[0044] Furthermore, CPU10 can make the head unit 4 able to shake vertically and horizontally, can move tail 5A of the tail section unit 5, or makes it act by making the required actuators 251 and 252 and ... drive based on the decision

result concerned to make it walk by making each leg unit 3A-3D drive etc. while it opts for the action which continues based on the control program stored in this decision result and DRAM11.

[0045] Moreover, in this case, by generating voice data if needed and giving this to a loudspeaker 24 as a sound signal through a digital disposal circuit 14, CPU10 makes the voice based on the sound signal concerned output outside, or turns on or blinks above-mentioned LED.

[0046] Thus, in this robot equipment 1, it is made as [ act / according to the situation of self and a perimeter the command from a user, and influence / it / autonomously ].

[0047] (2) the software configuration of a control program -- here comes to show the software configuration of the above-mentioned control program in robot equipment 1 to drawing 3 . In this drawing 3 , the device driver layer 30 is located in the lowest layer of this control program, and consists of device driver sets 31 which consist of two or more device drivers. In this case, each device driver is the object allowed to carry out direct access to the hardware used by CCD camera 20 ( drawing 2 ) and usual computers, such as a timer, and processes in response to interruption from corresponding hardware.

[0048] Moreover, the ROBOTIKKU server object 32 With the virtual robot 33 which becomes by the software group which offers the interface for being

located in the lowest layer of the device driver layer 30, for example, accessing hardware, such as above-mentioned various sensors and actuators 251 and 252, and ... With the power manager 34 who becomes by the software group which manages a switch of a power source etc. It consists of a device driver manager 35 who becomes by the software group which manages other various device drivers, and a dither INDO robot 36 which becomes by the software group which manages the device of robot equipment 1.

[0049] The manager object 37 consists of an object manager 38 and a service manager 39. In this case, the object manager 38 is a software group which manages starting and termination of each software group contained in the ROBOTIKKU server object 32, the middleware layer 40, and the application layer 41, and a service manager 39 is a software group which manages connection of each object based on the initial entry between each object described by the connection file stored in the memory card 28 ( drawing 2 ).

[0050] The middleware layer 40 is located in the upper layer of the ROBOTIKKU server object 32, and consists of software groups which offer the fundamental function of these robot equipments 1, such as an image processing and speech processing. Moreover, the application layer 41 is located in the upper layer of the middleware layer 40, and consists of software groups for opting for action of robot equipment 1 based on the processing result processed by each software

group which constitutes the middleware layer 40 concerned.

[0051] In addition, the concrete software configuration of the middleware layer 40 and the application layer 41 is shown in drawing 4 .

[0052] As shown in drawing 4 , in the middleware layer 40 For noise detection, The object for temperature detection, the object for brightness detection, the object for scale recognition, the object for distance detection, for posture detection, The recognition system 60 which has the input semantics converter module 59 etc. in the object for touch sensors, the object for motion detection, and each signal conditioning module 50 for color recognition - 58 lists, It consists of output systems 69 which have each signal conditioning modules 61-67 for the object for posture management, the object for tracking, the object for motion playback, the object for a walk, the object for a fall return, the object for LED lighting, and sound playback etc. in output semantics converter module 68 list.

[0053] Each signal conditioning modules 50-58 of the recognition system 60 incorporate the data with which it corresponds of each sensor data read from DRAM11 ( drawing 2 ) by the virtual robot 33 of the ROBOTIKKU server object 32, or image data and voice data, perform predetermined processing based on the data concerned, and give a processing result to the input semantics converter module 59. Here, the virtual robot 33 is constituted by the predetermined protocol as a part which carries out transfer or conversion of a

signal.

[0054] The input semantics converter module 59 It is based on the processing result given from these signal conditioning modules 50-58. "The fall was detected", [ "it is "noisy", hot / "hot" /, and bright", "the ball having been detected", and ] The self of "it was stroked", "it having been struck", "the scale of C-E-G having been heard", "the body which moves having been detected", "having detected the obstruction", etc. and a surrounding situation, the command from a user, and influence are recognized, and a recognition result is outputted to the application layer 41 ( drawing 3 ).

[0055] The application layer 41 consists of five modules, the behavioral model library 70, the action change module 71, the study module 72, the feeling model 73, and the instinct model 74, as shown in drawing 5 .

[0056] As shown in drawing 6 , "when a dc-battery residue decreases, and carrying out a fall return, and expressing [ avoiding an obstruction, and ] feeling", the behavioral models 701-70n which became independent, respectively "are formed [ the behavioral model library 70 is made to correspond to the condition item of the shoes "at the time of detecting a ball" etc. chosen beforehand, respectively, and ]". [ "being formed in it" ] [ "being formed" ]

[0057] And these behavioral models 701-70n The time of a recognition result being given from the input semantics converter module 59, respectively, The

parameter value of the corresponding emotion currently held like the after-mentioned at the feeling model 73 if needed when fixed time amount has passed, after the last recognition result is given, It opts for the action which continues while referring to the parameter value of the corresponding desire currently held at the instinct model 74, respectively, and a decision result is outputted to the action change module 71.

[0058] In the case of the gestalt of this operation, in addition, each behavioral models 701-70n As the technique of opting for the next action As opposed to the arc ARC1 - ARCn which connect between each node NODE0 - NODEn(s) for to other nodes NODE0 of which - NODEn it changes from one node (condition) NODE0 as shown in drawing 7 - NODEn The algorithm called the finite stochastic automaton determined probable based on the transition probability P1-Pn by which a \*\*\*\*\* setup was carried out is used.

[0059] Concretely, each behavioral models 701-70n are made to correspond to the node NODE0 which forms the self behavioral models 701-70n, respectively - NODEn, respectively, and have the state transition table 80 as shown in drawing 8 for every these node NODE0 - NODEn.

[0060] In this state transition table 80, the input event (recognition result) made into transition conditions in that node is listed by the train of an "input event name" at a priority, and the further conditions about that transition condition are



described by the corresponding line in the train of a "data name" and the "data range."

[0061] therefore, in the node NODE100 expressed in the state transition table 80 of drawing 8 When the recognition result of "detecting a ball (BALL)" is given The range of "magnitude (SIZE)" of the ball given with the recognition result concerned is "0 to 1000", When the recognition result of "detecting an obstruction (OBSTACLE)" is given, they have been conditions for that the range of "the distance (DISTANCE)" to the obstruction done with the recognition result concerned is "0 to 100" to change to other nodes.

[0062] Moreover, in this node NODE100, when there is no input of a recognition result, it also sets. The inside of each emotion held at the feeling model 73 and the instinct model 74 which 701-70n of behavioral models refers to periodically, respectively, and the parameter value of each desire, it was held at the feeling model 73 -- "-- glad (JOY) -- " -- "-- surprised (SURPRISE) -- " -- or -- "-- feeling sad (SADNESS) -- " -- when the range of one of parameter value is "50 to 100", it can change to other nodes.

[0063] moreover -- a state transition table 80 -- "-- others, while the node name which can change from the node NODE0 - NODEn in the line of the "transition place node" in the column of transition probability" to a node is listed It is described by the part where it corresponds in the column of transition probability"

to a node, respectively. the transition probability to each of other node NODE0 which can change when all the conditions described by the "input event name", the "data value", and the train of "the range of data" are met - NODEn -- "-- others -- the action which should be outputted in case it changes to the node NODE0 - NODEn -- "-- others -- it is described by the line of "output action" in the column of transition probability" to a node. in addition -- "-- others -- the sum of the probability of each line in the column of transition probability" to a node is 100 [%].

[0064] therefore, in the node NODE100 expressed in the state transition table 80 of drawing 8 for example, when the recognition result that it carries out "detecting a ball (BALL)", and the range of "magnitude (SIZE)" of the ball is "0 to 1000" is given It can change to "a node NODE120 (node 120)" by the probability of "30 [%]", and action of "ACTION 1" will be then outputted.

[0065] When they are constituted as a lot of nodes NODE0 described as such [ respectively ] a state transition table 80 - NODEn(s) are connected, and a recognition result is given from the input semantics converter module 59, each behavioral models 701-70n opt for the next action probable using the state transition table 80 of the node NODE0 - NODEn, and are made as [ output / to the action change module 71 / a decision result ].

[0066] The action change module 71 shown in drawing 5 chooses from each

behavioral models 701-70n of the behavioral model library 70 the action outputted from the high behavioral models 701-70n of the priority beforehand defined among the actions outputted, respectively, and sends out the command (this is hereafter called action command.) of the purport which should perform the action concerned to the output semantics converter module 68 of the middleware layer 40. In addition, in the gestalt of this operation, priority is set up highly about 701-70n of behavioral models written by the bottom in drawing 6 .

[0067] Moreover, the action change module 71 notifies that the action was completed based on the completion information of action given from the output semantics converter module 68 after the completion of action to the study module 72, the feeling model 73, and the instinct model 74.

[0068] On the other hand, the recognition result of the instruction received as influence from a user, such as the study module 72 "was struck" among the recognition results given from the input semantics converter module 59 and "it having been stroked", is inputted.

[0069] And based on the notice from this recognition result and the action change module 71, the study module 72 reduces the manifestation probability of that action, when "struck" (scolded), and when "stroked" (praised), it changes behavioral models [ in the behavioral model library 70 / 701-70n ] corresponding transition probability so that the manifestation probability of that action may be

raised.

[0070] on the other hand, the feeling model 73 -- "-- glad (joy) -- " -- "-- feeling sad (sadness) -- " -- "-- getting angry (anger) -- " -- "-- surprised (surprise) -- " -- "dislike (disgust)" -- and -- "-- afraid (fear) -- " -- the parameter with which the strength of the emotion is expressed for every emotion is held about a total of six emotions. And the feeling model 73 updates the parameter value of each [ these ] emotion periodically based on the notice from the specific recognition result to which it is given from the input semantics converter module 59, respectively, such as "it having been struck" and "it having been stroked", and elapsed time and the action change module 71 etc.

[0071] The recognition result to which the feeling model 73 is specifically given from the input semantics converter module 59, The variation of the action and its emotion when being computed by predetermined operation expression based on the elapsed time after updating last time etc. of the robot equipment 1 at that time  $\Delta E [t]$ , The multiplier which expresses the sensibility of  $E [t]$  and its emotion for the parameter value of the current emotion is set to  $k_e$ . (1) By the formula, as parameter value [ of the emotion in the following period ]  $E [t+1]$  is computed and this is replaced with parameter value [ of the current emotion ]  $E [t]$ , the parameter value of the emotion is updated. Moreover, the feeling model 73 updates the parameter value of all emotions like this.

[0072]

[Equation 1]

$$E[t+1]=E[t]+k_e \times \Delta E[t] \quad \dots (1)$$

[0073] In addition, it is decided beforehand how many effects the notice from each recognition result or the output semantics converter module 68 has on amount of fluctuation  $\Delta E[t]$  of the parameter value of each emotion  $E[t]$ . For example, the recognition result of "having been struck" has big effect on amount of fluctuation  $\Delta E[t]$  of the parameter value of the emotion of the "resentment"  $E[t]$ , and the recognition result of "having been stroked" has big effect on amount of fluctuation  $\Delta E[t]$  of the parameter value of the emotion of "joy"  $E[t]$ .

[0074] Here, the notice from the output semantics converter module 68 is the so-called feedback information (the completion information of action) of action, and is the information on the appearance result of action, and the feeling model 73 changes feeling also using such information. this -- for example, the feeling level of the resentment falls by action of "barking" -- like -- they are things. In addition, the notice from the output semantics converter module 68 is inputted also into the study module 72 mentioned above, and the study module 72 changes behavioral models [ 701-70n ] corresponding transition probability based on the notice.

[0075] In addition, feedback of an action result may be made with the output (action to which feeling was added) of the action change module 71.

[0076] on the other hand, "movement avarice (exercize)", "love avarice (affection)", "appetite (appetite)", and "the curiosity (curiosity) of the instinct model 74" are mutually-independent -- the parameter with which the strength of the desire is expressed for these the desires of every is held about four desires the bottom. And the instinct model 74 updates the parameter value of these desires periodically based on the recognition result to which it is given from the input semantics converter module 59, respectively, the notice from elapsed time and the action change module 71, etc.

[0077] The instinct model 74 specifically about "movement avarice", "love avarice", and "curiosity" The amount of fluctuation of that the desire of the when being computed by predetermined operation expression based on the notice from a recognition result, elapsed time, and the output semantics converter module 68 etc.  $\Delta I[k]$ , The multiplier which expresses the sensibility of  $I[k]$  and its desire for the parameter value of the current desire is set to  $k_i$ . As parameter value [ of that desire in the following period ]  $I[k+1]$  is computed using (2) types with a predetermined period and this result of an operation is replaced with parameter value [ of that current desire ]  $I[k]$ , the parameter value of that desire is updated. Moreover, the instinct model 74 updates the parameter value

of each desire except "appetite" like this.

[0078]

[Equation 2]

$$I[k+1]=I[k]+k_i \times \Delta I[k] \quad \dots (2)$$

[0079] In addition, it is decided beforehand how many effects the notice from a recognition result and the output semantics converter module 68 etc. has on amount of fluctuation  $\Delta I[k]$  of the parameter value of each desire  $I[k]$ , for example, it has effect to amount of fluctuation  $\Delta I[k]$  of the parameter value of the "fatigue"  $I[k]$  with the big notice from the output semantics converter module 68.

[0080] In addition, in the gestalt of this operation, it is regulated so that each emotion and the parameter value of each desire (instinct) may be changed in the range from 0 to 100, respectively, and the value of multipliers  $k_e$  and  $k_i$  is also set up according to the individual for each [ an emotion and ] the desire of every.

[0081] On the other hand, abstract action commands, such as it being [ which is given from the action change module 71 of the application layer 41 as mentioned above ] "advance", "it being glad", the output semantics converter module 68 of the middleware layer 40 "cries", as shown in drawing 4 , or "tracking (a ball is pursued)", are given to the signal conditioning modules 61-67 with which the

output system 69 corresponds.

[0082] And these signal conditioning modules 61-67 The servo command value which should be given to the actuators 251-25n ( drawing 2 ) in order to perform this action based on the action command concerned, if an action command is given, The drive data given to the voice data of the sound outputted from a loudspeaker 24 ( drawing 2 ) and/or LED of a "eye" are generated. Sequential sending out of these data is carried out at the actuators 251-25n which correspond through the virtual robot 33 and digital disposal circuit 14 ( drawing 2 ) of the ROBOTIKKU server object 32 one by one, a loudspeaker 24, or LED.

[0083] Thus, in robot equipment 1, it is made based on the control program as [ perform / the situation of self (interior) and a perimeter (exterior), the directions from a user, and autonomous action according to influence ].

[0084] (3) Use the approach called shaping (Shaping) to robot equipment 1 as an approach of teaching action with the gestalt of the action instruction approach book operation used with the gestalt of this operation. By this approach, in order to form action, action is divided into the continuous small response which can be attained, and, finally it leads to desired action. That is, this approach is gradually guided towards suitable action. Although various techniques can be used for performing each phase, with the gestalt of this operation, the approach called clicker training which has spread as the training approach of an animal is used.



Hereafter, this clicker training is explained.

[0085] Clicker training is based on the theory of the operant conditioning of B.F. Skinner indicated by "B. F. Skinner, "The Behavior of Organisms., Appleton Century Crofts, New York, NY., 1938." Karen Pryor which is a behaviorist psychology person's Gary Wilkes and the trainer of a dolphin spread the training approach of a dog based on this clicker training through the 1980s together. By clicker training, the clicker which is a metal small toy is used instead of the whistle currently conventionally used for training of a dolphin.

[0086] A push on a clicker emits a short sharp sound. Any semantics cannot be found for an animal only to this sound. However, a trainer relates this sound with primary reinforcer. An animal senses primary reinforcer instinctive as remuneration like food or a toy. The sound of a clicker serves as a secondary reinforcement factor (conditioned reinforcer) by relating primary reinforcer and many times and carrying out. This 2nd reinforcer carries out the role of the positive hint that remuneration can be got soon, for an animal. The clicker itself is used in order to lead not remuneration but an animal to desired action for an animal. Moreover, an animal can be told about which action should be strengthened with this clicker. A trainer gives primary reinforcer, only when an animal carries out desired action. This is the signal showing termination of a series of induction processes led to desired action.

[0087] Clicker training consists of four steps shown below at least.

[0088] The 1st step: Carry out the charge up of the clicker. In this step, an animal has prizes (food etc.) and chain attachment \*\*\*\*\* in the sound of a clicker. In it, a clicker is sounded, and continuously, it repeats giving remuneration continuously about many times, for example, 20 to 50 times, until an animal comes to be clearly excited to the sound of a clicker.

[0089] The 2nd step: Catch action. Next, it guides so that actuation of a request of an animal may be taken. For example, when it is thought that I want an animal for a trainer to draw and turn around a circle in the clockwise direction, and the head of an animal moves to the right a little, it begins from sounding a clicker. If an animal comes to move the head repeatedly, a clicker will be sounded only when beginning to move the body to the right this time. Remuneration will be given, if the criteria with which a clicker is sounded gradually are raised and the body rotates one time completely.

[0090] The 3rd step: Give a command word. If an animal learns desired action, a command word will be said only once. The timing which says a command word is just before immediately after an animal performs the action.

[0091] The 4th step: Try action. Next, it is necessary to try and refine the learned action. Only when action with an exact animal is taken, a command word is said and remuneration is given.

[0092] As explained above, clicker training is for leading an animal to desired action. Furthermore, it can be used that an animal performs this clicker training spontaneously in order to make the rare action which is not rash learn, or in order to make a series of continuous actions learn.

[0093] Below, how the robot equipment 1 mentioned above by using the principle of this clicker training learns action is explained.

[0094] (4) Explain the important section to which this invention was applied in the application above-mentioned robot equipment 1 to robot equipment. Robot equipment 1 is related with a picture signal or a sound signal (acoustic signal), action is learned, and after study is constituted so that the picture signal and sound signal which were associated may cause action. Although the following explanation mainly explains the example which relates voice with action as study, it cannot be overemphasized that an image can be related with action, either. Specifically, robot equipment 1 has the following configurations for implementation of this invention.

[0095] Robot equipment 1 is equipped with the input section 101, the primary reinforcer detecting element 102, the secondary reinforcement factor associative section 103, the action decision section 104, the action storage section 105, the action output section 106, the speech recognition section 107, and the associative storage section 108 as shown in drawing 9 .

[0096] Here, the input section 101 is the microphone 23 and touch sensor 21 which are shown in drawing 2 . The input section 101 supplies a sound signal or a sensor signal to the primary reinforcer detecting element 102 and the secondary reinforcement factor associative section 103. Moreover, the input section 101 supplies a sound signal to the speech recognition section 107.

[0097] The primary reinforcer detecting element 102 detects primary reinforcer from the sound signal supplied from the input section 101, or a sensor signal. Primary reinforcer is set up so that robot equipment 1 may sense as remuneration beforehand, and detection of the pressure (what depends the head on striking lightly) to the head of robot equipment 1, the sound signal with which it loves and \*\*\*\* is expressed are mentioned. Below, the sound signal of "bravo" is used as primary reinforcer. The primary reinforcer detecting element 102 will supply primary strengthening signals to the action decision section 104 and the action storage section 105, if primary reinforcer is detected.

[0098] The secondary reinforcement factor associative section 103 reminds of it and detects a secondary reinforcement factor from the sound signal supplied from the input section 101, or a sensor signal. The secondary reinforcement factor may be beforehand set as robot equipment 1 like primary reinforcer, and a trainer relates it with primary reinforcer, and you may make it set it up. When a trainer sets up a secondary reinforcement factor, robot equipment 1 is made into

secondary reinforcement factor setting mode, and it sets up by [ as being shown in drawing 10 ].

[0099] As shown in drawing 10 , in case a secondary reinforcement factor is set up, in step S1, the stimulus used as a secondary reinforcement factor is given first. Even if this stimulus is a contact stimulus and it is an acoustic stimulus, it is not cared about.

[0100] Then, in step S2, the stimulus which gave primary reinforcer and was given at step S1 in step S3 is related with primary reinforcer, and it memorizes as a secondary reinforcement factor.

[0101] In addition, if it detects, for example that primary reinforcer was given after the stimulus which is not limited to this example, for example, has robot equipment 1 30 times or more many times, the setting technique of a secondary reinforcement factor may be beforehand set up so that that stimulus may be memorized as a secondary reinforcement factor. Moreover, modification of the set-up secondary reinforcement factor can also be enabled. A secondary reinforcement factor can be changed by giving a stimulus of arbitration, giving in this case, for example, a secondary reinforcement factor change command.

[0102] In addition, this secondary reinforcement factor can be given quickly, and it must be easy to detect it so that it can function as an indicator which excelled for guiding robot equipment 1 to suitable action. Below, the sound signal of

"good" is used as a secondary reinforcement factor.

[0103] The secondary reinforcement factor associative section 103 will supply a secondary reinforcement signal to the action decision section 104 and the action storage section 105, if a secondary reinforcement factor is detected.

[0104] The action decision section 104 opts for the next action probable based on the transition probability to other nodes [ node / (condition) / a certain ], and supplies a decision result to the action output section 106. Moreover, the action decision section 104 makes the action storage section 105 supply and memorize the hysteresis of the action for which it opted. Furthermore, based on primary strengthening signals and secondary reinforcement signal which were supplied from the primary reinforcer detecting element 102 and the secondary reinforcement factor associative section 103, the action decision section 104 updates transition probability, or it updates transition probability so that it may be easy to change to the node (condition) and may become it from the last node (condition), so that the selection probability of a certain node (condition) may become high, so that it may mention later.

[0105] The action storage section 105 memorizes the hysteresis of the action for which it opted in the action decision section 104. Moreover, to the action storage section 104, primary strengthening signals and secondary reinforcement signal from the primary reinforcer detecting element 102 and the secondary

reinforcement factor associative section 103 are supplied, and the action to which the secondary reinforcement signal was given after becoming action learning mode before primary strengthening signals were supplied is memorized as a series of actions in it.

[0106] The action output section 106 is sent out to the output semantics converter module 68 which shows the action command mentioned above to drawing 4 based on the decision result of the action supplied from the action decision section 105.

[0107] The speech recognition section 107 carries out speech processing of the sound signal inputted from the input section 101, and recognizes it as predetermined language. HMM (HiddenMarkov Model) is used for the speech recognition section 107, it is constituted, and, specifically, recognizes the inputted voice as a phoneme sequence by two or more recognition classes depended on HMM.

[0108] The associative storage section 108 associates and carries out associative storage of the command showing a series of actions to which the secondary reinforcement signal was given after becoming action learning mode before primary strengthening signals were supplied, and the action concerned recognized in the speech recognition section 107. A series of actions by which associative storage was carried out by relating with a command are memorized

by the action storage section 104.

[0109] A trainer can guide robot equipment 1 to desired action using a secondary reinforcement factor because robot equipment 1 has the above configurations. That is, a trainer will give a secondary reinforcement factor, if robot equipment 1 approaches desired action, and he is taken as the indicator for approaching the action. If robot equipment 1 finally acts a request, a trainer will give primary reinforcer and robot equipment will memorize a series of actions to which the secondary reinforcement factor till then was given. And if a trainer utters a command, robot equipment will output a series of memorized actuation in order. In addition, it is also possible to narrow them down, as a series of memorized actions are mentioned later.

[0110] A series of procedures of this action study approach are shown in drawing 11 . As shown in drawing 11 , in step S10, it is first set as action learning mode.

[0111] Next, in step S11, the transition probability from a node to a node is initialized, and in step S12, robot equipment 1 opts for action according to current transition probability, and outputs the action.

[0112] At step S13, it is distinguished whether primary reinforcer was given by action outputted at step S12. When primary reinforcer is given, it progresses to step S17. In step S13, when primary reinforcer is not given, it progresses to step



S14.

[0113] At step S14, it is distinguished whether the secondary reinforcement factor was given by action outputted at step S12. When the secondary reinforcement factor is not given, it supposes that it is not action for approaching desired action, it returns to step S12, and opts for the next action again. In step S14, when a secondary reinforcement factor is given, it progresses to step S15.

[0114] At step S15, it memorizes as action which results in action of a request of the action outputted at step S12, and progresses to step S16.

[0115] At step S16, transition probability is updated and it returns to step S12 so that the selection probability of action for the secondary reinforcement signal to have been given may become high.

[0116] At step S17, the command for relating with the signal with which primary strengthening signals were given is uttered.

[0117] At step S18, the action to which the secondary reinforcement signal was given by the time it resulted in the action to which primary strengthening signals were given, and its action is related with the command uttered at step S17 as a series of actions, and it memorizes and ends.

[0118] By the action study approach in the gestalt of this operation, robot equipment 1 is guided to desired action by the above procedures, and the action is learned.

[0119] In addition, although the command related with desired action was uttered in the flow chart of drawing 11 after initializing transition probability, you may make it utter the command related with the action immediately after it is not limited to this and robot equipment 1 acts a request.

[0120] Moreover, although the flow chart of drawing 11 explained as what updates transition probability so that the selection probability of action for the secondary reinforcement signal to have been given might become high in step S16, it is not limited to this and you may make it update transition probability so that it may be easy to change to the action to which the secondary reinforcement signal was given from the last action and may become it.

[0121] By the way, although a trainer can guide robot equipment 1 to desired actuation using a secondary reinforcement factor, he needs to define the "phase" of action of robot equipment 1 as this first. Namely, after robot equipment 1 takes a certain action, it and the change to "similar" action must be possible for it. As an approach for defining the phase of existing action, two examples are given to below.

[0122] The 1st approach explains each action using 1 set of properties according to which it is classified into an explanatory property and an intention-property first.

[0123] An explanatory property is for example, an initiation posture (it stands,

sits down and lies), and whether a robot's head, a foot, other parts, or a sound is emitted actually relates to this.

[0124] An intention-property is the purpose which makes action take, for example, the thing for moving, or expresses whether it is for obtaining whether it is and the attention for holding something. Each action can be seen as a point in the space defined by these dimensions. All actions can be format-ized according to this property, and it can guide to the action "was similar" from a certain action by defining the "distance" during two actions.

[0125] Although all actions can be considerably expressed by this approach in a short time once a property is chosen, it is hard to predict the transition during action.

[0126] Next, the 2nd approach creates the graph which shows the transition during each action. According to this approach, transition of the action which robot equipment 1 takes can be grasped more effectively. It is possible like the 1st approach to combine the objective similar point during action with a certain criteria relevant to "intention nature." Furthermore, general action (it stands [ sitting down, ]) and rare action (the gymnastics which dance a special dance are carried out) can be distinguished more clearly. A part of phase defined as drawing 12 in the graph former rhythm is shown. Moreover, a part of condition of having taken the action to which robot equipment 1 was mentioned to drawing

12 is shown in drawing 13 .

[0127] A trainer explains concretely the case where the language for the excavation action whose robot equipment 1 is the rarely taken special action is taught, using drawing 12 . In this action, as shown in drawing 13 (G), a forward left foot is used for robot equipment 1 sitting down and scratching the ground. Moreover, the head is looking at the foot and follows the motion. This action corresponds to the condition of the node [DIG] of drawing 12 .

[0128] First, robot equipment 1 presupposes that it has started in the condition of a node [STAND] as shown in drawing 13 (A). As first shown in drawing 13 (B), robot equipment 1 begins a walk, namely, changes in the condition of a node [WALK] from the condition of a node [STAND]. Since the transition to this node [WALK] does not approach the node [DIG] which is in a desired condition, a trainer gives no reinforcer. Then, robot equipment 1 starts another action from the condition of a node [STAND]. In addition, you may make it return to the condition of a node [STAND] automatically, and may make it return according to transition probability in this case.

[0129] Next, robot equipment 1 sits down namely, changes to a node [SIT], as shown in drawing 13 (D). Since the transition to this node [SIT] is a thing approaching the node [DIG] which is in a desired condition, a trainer says "good" and gives a secondary reinforcement factor to robot equipment 1. Transition

probability is updated by this secondary reinforcement factor being given, and robot equipment 1 tends to take the action (node [SIT]) which is the action to which the secondary reinforcement factor was given and which sits down, and the associated action.

[0130] Then, as shown in drawing 13 (E), it is begun to push robot equipment 1 by two nose gear. That is, it changes to a node [PUSH]. However, since the transition to this node [PUSH] does not approach the node [DIG] which is in a desired condition, a trainer gives no reinforcer. Then, robot equipment 1 starts another action from the condition of a node [SIT]. You may make it return to the condition of a node [SIT] automatically, and may make it return according to transition probability also in this case.

[0131] Then, robot equipment 1 changes to a node [HELLO], as shown in drawing 13 (F), raises a forward left foot and operates a "halo." Since the forward left foot which uses this action for excavation is related, a trainer tells "good" to robot equipment 1 and gives a secondary reinforcement factor.

[0132] If robot equipment 1 changes to a node [DIG] and starts excavation action through the exchange with the above trainers and robot equipment 1, a trainer will say "bravo" and will give primary reinforcer to robot equipment 1. As for this primary reinforcer, robot equipment 1 means having acted the request. Robot equipment 1 memorizes a series of actions called [SIT]-[HELLO]-[DIG] the

above result.

[0133] In addition, even if it uses not only the path of [SIT]-[HELLO]-[DIG] mentioned above but another path, robot equipment 1 can be guided to desired action. For example, a trainer can move crus sinistram diaphragmatis to robot equipment 1 through the action to kick as shown in drawing 13 (C), i.e., a node, [KICK], and he can guide so that it may sit on a degree.

[0134] If desired action is performed and primary reinforcer is given, a trainer will give the command word immediately related with action to robot equipment 1. In order to check that surely robot equipment 1 has understood the command word in that case, it is desirable to adopt a certain feedback. As feedback, robot equipment 1 imitates the command through the loudspeaker 24 shown in drawing 2 , for example, and you may make it require a check, and the purport that he understood something by blinking LED which is a "eye" etc. is shown, and you may make it wait to repeat a command again repeatedly. When robot equipment 1 cannot detect the command again given by the trainer, you may make it the special action (for example, for the head to be shaken) for doing the thing same once again express that he was not able to understand a command to a trainer.

[0135] Robot equipment 1 performs a series of actions called [SIT]-[HELLO]-[DIG] of a memorized single string acted for example, mentioned

above, when this command word and a series of actions mentioned above are associated and memorized and a command word is given again.

[0136] When a trainer wants to associate the excavation action of a series of actions, and a command word here, it is necessary to narrow down the action of a series of. For this reason, if a command word is given, robot equipment 1 can repeat a series of memorized actuation spontaneously. When primary reinforcer is given after robot equipment 1 performed a series of actuation again, it is judged that robot equipment 1 is connected with action of a single string [ command / the ]. When that is not right, although it originates in a series of actions performed before, rather than it, robot equipment 1 makes little the new actions of the number of phases of a series of, and performs the new action of a series of. Thus, finally the command word can judge that robot equipment 1 relates only to excavation action.

[0137] As explained above, the robot equipment 1 in the gestalt of this operation can learn the action efficiently by teaching action like the time of a trainer teaching an animal and a man action. Moreover, since it is guided to desired action by using a secondary reinforcement factor as an indicator in that case, the count of the trial-and-error when learning complicated actuation and the action which is not performed rarely can be reduced. Furthermore, when a secondary reinforcement factor is given, by updating transition probability, the probability

which changes to desired action becomes high, and becomes possible [ reaching early by desired action ].

[0138] By such a function being included in robot equipment 1, communication with natural user (owner) and robot equipment 1 can be aimed at. In addition, this control program is stored in the memory card 28 or flash ROM 12 of drawing 2.

[0139] In addition, as for this invention, it is needless to say for various modification to be possible in the range which is not limited only to the gestalt of operation mentioned above and does not deviate from the summary of this invention.

[0140] For example, although speech information was used as a command related with action with the gestalt of operation mentioned above, you may be the image information through the contact information which is not limited to this and minded the touch sensor, or CCD.

[0141] Moreover, although clicker training was used with the gestalt of operation mentioned above when teaching action to robot equipment, it is not limited to this and can use for other purposes. That is, when teaching an objective identifier, robot equipment is guided to the object by giving a secondary reinforcement factor, in case robot equipment approaches an object. Thereby, the difficult problem of joint cautions is mitigable.



[0142] Moreover, you may make it combine clicker training and the lure ring mentioned above. It becomes unnecessary thereby, to wait to perform action of a certain phase spontaneously.

[0143]

[Effect of the Invention] The robot equipment applied to this invention as explained to the detail above An action decision means to be robot equipment which a right hand side is controlled [ robot ] and makes action appear, and to opt for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which the above-mentioned action decision means opted. An action output means to make the action concerned for which it opted appear, and an input means to input external information, A primary reinforcer detection means to detect the primary reinforcer beforehand set up from the above-mentioned external information which the above-mentioned input means inputted, A secondary reinforcement factor detection means to detect a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has an action storage means to memorize the action for which the above-mentioned action decision means opted. The above-mentioned action decision means Whenever the above-mentioned secondary reinforcement factor is detected, the appearance inclination of action for the secondary reinforcement

factor concerned to have been detected is raised, and the above-mentioned action storage means is characterized by the primary reinforcer concerned memorizing the action by which it was detected at least, if the above-mentioned primary reinforcer is detected.

[0144] Here, the action by which the secondary reinforcement factor was detected is action until it results in the action by which primary reinforcer was detected.

[0145] Such robot equipment will memorize the action by which the primary reinforcer concerned was detected at least, if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it, whenever a secondary reinforcement factor is detected.

[0146] The robot equipment with which this invention was applied can be given to the action which is going to learn a secondary reinforcement factor as an indicator, and can learn action efficiently.

[0147] In order to attain the purpose mentioned above, moreover, the action study approach of the robot equipment concerning this invention The action decision process of being the action study approach of the robot equipment which a right hand side is controlled [ robot ] and makes action appearing, and opting for action of 1 based on the appearance inclination of action, The

above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process If the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised and the above-mentioned primary reinforcer is detected at the above-mentioned action storage process whenever the above-mentioned secondary reinforcement factor is detected, it is characterized by memorizing the action by which the primary reinforcer concerned was detected at least.

[0148] Here, the action by which the secondary reinforcement factor was detected is action until it results in the action by which primary reinforcer was detected.

[0149] By such action study approach of robot equipment, if the appearance

inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it whenever a secondary reinforcement factor is detected, the action by which the primary reinforcer concerned was detected at least will be memorized.

[0150] The robot equipment with which this invention was applied can be given to the action which is going to learn a secondary reinforcement factor as an indicator, and can learn action efficiently.

[0151] In order to attain the purpose mentioned above, moreover, the action learning program of the robot equipment concerning this invention The action decision process of being the action learning program of robot equipment which a right hand side is controlled [ learning program ] and makes action appearing, and opting for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the

above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process If the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised and the above-mentioned primary reinforcer is detected at the above-mentioned action storage process whenever the above-mentioned secondary reinforcement factor is detected, it is characterized by memorizing the action by which the primary reinforcer concerned was detected at least.

[0152] Here, the action by which the secondary reinforcement factor was detected is action until it results in the action by which primary reinforcer was detected.

[0153] In such an action learning program of robot equipment, if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it whenever a secondary reinforcement factor is detected, the action by which the primary reinforcer concerned was detected at least will be memorized.

[0154] The robot equipment with which this invention was applied can be given to the action which is going to learn a secondary reinforcement factor as an

indicator, and can learn action efficiently.

[0155] In order to attain the purpose mentioned above, moreover, the program documentation medium concerning this invention It is the program documentation medium by which the action learning program of robot equipment which a right hand side is controlled [ learning program ] and makes action appear was recorded and in which computer control is possible. The action decision process that the action learning program of the above-mentioned robot equipment opts for action of 1 based on the appearance inclination of action, The above-mentioned right hand side is controlled based on the action for which it opted at the above-mentioned action decision process. The primary reinforcer detection process of detecting the primary reinforcer beforehand set up from the above-mentioned external information which an input means to input external information as the action output process of making the action concerned for which it opted appearing inputted, The secondary reinforcement factor detection process of detecting a secondary reinforcement factor from the above-mentioned external information which the above-mentioned input means inputted, It has the action storage process of memorizing the action for which it opted at the above-mentioned action decision process. At the above-mentioned action decision process Whenever the above-mentioned secondary reinforcement factor is detected, the appearance inclination of action for the

secondary reinforcement factor concerned to have been detected is raised. At the above-mentioned action storage process Detection of the above-mentioned primary reinforcer records the action learning program of the robot equipment characterized by memorizing the action by which the primary reinforcer concerned was detected at least.

[0156] Here, the action by which the secondary reinforcement factor was detected is action until it results in the action by which primary reinforcer was detected.

[0157] In the action learning program of the robot equipment currently recorded on such a program documentation medium, if the appearance inclination of action for the secondary reinforcement factor concerned to have been detected is raised at the time of study of action and primary reinforcer is detected at it whenever a secondary reinforcement factor is detected, the action by which the primary reinforcer concerned was detected at least will be memorized.

[0158] The robot equipment with which this invention was applied can be given to the action which is going to learn a secondary reinforcement factor as an indicator, and can learn action efficiently.

---

## DESCRIPTION OF DRAWINGS

---

[Brief Description of the Drawings]

[Drawing 1] It is the perspective view showing the appearance configuration of the robot equipment in the gestalt of operation of this invention.

[Drawing 2] It is the block diagram showing the circuitry of this robot equipment.

[Drawing 3] It is the block diagram showing the software configuration of this robot equipment.

[Drawing 4] It is the block diagram showing the configuration of the middleware layer in the software configuration of this robot equipment.

[Drawing 5] It is the block diagram showing the configuration of the application layer in the software configuration of this robot equipment.

[Drawing 6] It is the block diagram showing the configuration of the behavioral model library of this application layer.

[Drawing 7] It is drawing explaining the finite stochastic automaton used as the information for the action decision of this robot equipment.

[Drawing 8] It is drawing showing the state transition table prepared for each node of a finite stochastic automaton.

[Drawing 9] It is the block diagram showing the configuration of the part concerning this invention in this robot equipment.



[Drawing 10] It is a flow chart explaining the procedure in the case of setting up a secondary reinforcement factor.

[Drawing 11] It is a flow chart explaining a series of procedures of the action study approach of this robot equipment.

[Drawing 12] It is drawing explaining the transition during the action which this robot equipment performs.

[Drawing 13] It is drawing explaining a part of action which this robot equipment performs.

[Drawing 14] It is drawing explaining the example of the action instruction approach for the conventional animal.

[Description of Notations]

1 Robot Equipment, 101 Input Section, 102 Primary Reinforcer Detecting Element, 103 2nd Reinforcer Associative Section, 104 Action Decision Section, 105 Action Storage Section, 106 Action Output Section, 107 Speech Recognition Section, 108 Associative Storage Section